



UNIWERSYTET ŚLĄSKI
W KATOWICACH

Wydział Humanistyczny
Instytut Językoznawstwa

mgr Aleksandra Zięba
nr albumu: 9027

**Wyodrębnianie polskich, rosyjskich i czeskich
jednostek wielowyrazowych o funkcji fatycznej
i emotywniej z użyciem narzędzi i zasobów
korpusowych. Studium leksykograficzne**

Rozprawa doktorska
napisana pod kierunkiem
prof. dr hab. Andrzeja Charciarka

Promotor pomocniczy
dr Maciej Labocha

Sosnowiec 2025

Słowa kluczowe: *reprodukt, leksykografia przekładowa, lingwistyka korpusowa, frazematyka, wielowyrazowa jednostka języka, fatyzm, emotywność*

Oświadczenie autora pracy

Ja, niżej podpisana:

imię (imiona) i nazwisko: Aleksandra Jagoda Zięba

autor pracy dyplomowej pt.

Numer albumu: 9027

Doktorantka Wydziału Humanistycznego Uniwersytetu Śląskiego w Katowicach

dyscyplina: Językoznawstwo.

Oświadczam, że ww. rozprawa doktorska:

- została przygotowana przeze mnie samodzielnie¹,
- nie narusza praw autorskich w rozumieniu ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (tekst jednolity Dz. U. z 2006 r. Nr 90, poz. 631, z późn. zm.) oraz dóbr osobistych chronionych prawem cywilnym,
- nie zawiera danych i informacji, które uzyskałem w sposób niedozwolony,
- nie była podstawą nadania dyplomu uczelni wyższej lub tytułu zawodowego ani mnie, ani innej osobie.

Oświadczam również, że treść rozprawy doktorskiej zamieszczonej przeze mnie w Archiwum Prac Dyplomowych jest identyczna z treścią zawartą w wydrukowanej wersji pracy.

Jestem świadoma odpowiedzialności karnej za złożenie fałszywego oświadczenia.

.....
Data

.....
Podpis autora pracy

¹ Uwzględniając merytoryczny wkład promotorów (w ramach prowadzonego seminarium dyplomowego)

SPIS TREŚCI

WPROWADZENIE	6
ROZDZIAŁ I	
1. Leksykografia: zakres, definicje, podziały	12
1.1 Zarys historii i rozwoju leksykografii	16
1.2 Rozwój leksykografii w Polsce	18
1.3 Rozwój leksykografii w Rosji	20
1.4 Rozwój leksykografii w Czechach	23
1.5 Leksykografia jednojęzyczna wobec wielowyrazowych jednostek języka mówionego	26
1.6 Rejestracja jednostek fatycznych i emotywnych we współczesnych polskich słownikach jednojęzycznych	29
1.7 Rejestracja jednostek fatycznych i emotywnych we współczesnych rosyjskich słownikach jednojęzycznych	36
1.8 Rejestracja jednostek fatycznych i emotywnych we współczesnych czeskich słownikach jednojęzycznych	44
1.9 Leksykografia przekładowa wobec wielowyrazowych jednostek języka mówionego	50
1.10 Rejestracja jednostek fatycznych i emotywnych we współczesnych słownikach polsko-rosyjskich i rosyjsko-polskich	53
1.11 Rejestracja jednostek fatycznych i emotywnych we współczesnych słownikach polsko-czeskich i czesko-polskich	58
ROZDZIAŁ II	
2. Zjawisko reprodukowalności	65
2.1 Reprodukowalność w języku	66
2.2 Reprodukacja we frazeologii	71
2.3 Ogólne założenia frazematyki	76
2.4 Sytuacyjność we frazematyce	78
2.5 Osoba mówiąca i model aktu komunikacji	82
2.6 Status semantyczny frazemów	84
2.7 Reprodukowalność we frazematyce	87
ROZDZIAŁ III	
3. Lingwistyka korpusowa. Rys historyczny	91
3.1 Korpus językowy	95
3.2 Typy korpusów	97
3.3 Rodzaje analizy korpusowej	98
3.4 Podstawowe narzędzia badawcze korpusów	99
3.4.1 Wyszukiwanie interaktywne	99

3.4.2 Anotacja	99
3.4.3 Lematyzacja	101
3.4.4 Kolokator	102
3.4.5 Zastosowanie korpusów	103
3.5 Narodowy Korpus Języka Polskiego	104
3.5.1 Ogólna charakterystyka NKJP	104
3.5.2 Wyszukiwanie w NKJP	105
3.5.3 Kolokator w NKJP	109
3.5.4 Anotacja w NKJP	110
3.6 Korpus MoncoPL	111
3.6.1 Ogólna charakterystyka MoncoPL	111
3.6.2 Wyszukiwanie w MoncoPL	112
3.6.3 Fasety i trendy	114
3.6.4 Kolokacje	115
3.7 Narodowy Korpus Języka Rosyjskiego	116
3.7.1 Ogólna charakterystyka NKJR	116
3.7.2 Wyszukiwanie w NKJR	118
3.7.3 Rodzaje anotacji w NKJR	121
3.8 Narodowy Korpus Języka Czeskiego	122
3.8.1 Ogólna charakterystyka NKJCz	122
3.8.2 Wyszukiwanie w NKJCz	124
3.8.3 Anotacja w NKJCz	128
3.8.4 Kolokacje w NKJCz	130

ROZDZIAŁ IV

4. Korpusy wielojęzyczne	134
4.1 Korpus porównywalny	136
4.2 Korpus równoległy	138
4.3 Korpus Równoległy InterCorp	143
4.3.1 Ogólna charakterystyka InterCorp	143
4.3.2 Wyszukiwanie w InterCorp	145
4.3.3 Anotacja w InterCorp	149
4.3.4 Zalety i ograniczenia korpusu InterCorp	152
4.4 Korpus Równoległy Narodowego Korpusu Języka Rosyjskiego	154
4.4.1 Ogólna charakterystyka	154
4.4.2 Wyszukiwanie w korpusie równoległym i wielojęzycznym	157
4.4.3 Anotacja w Korpusie Równoległym NKJR	161
4.4.4 Zalety i ograniczenia korpusu Równoległego NKJR	164

ROZDZIAŁ V

5. Geneza pojęcia reproduktu: Inspiracje i rozwój myśli opolskiej	166
5.1 Reprodukt	170
5.2 Emotywy	176
5.3 Fatyzmy	183

ROZDZIAŁ VI

6. Metody wyodrębniania reproduktów fatycznych i emotywnych	192
6.1 Metoda słownikowo-kolokacyjna	199
6.2 Pomocniki formalne	227
6.3 Metaoperatory narracyjne	240
6.4 Metoda synonimiczna	258
6.5 Wyszukiwanie przyrostowe	283
6.6 Metoda tłumaczeniowa	302

ROZDZIAŁ VII

7. Metody ustalania ekwiwalentów tłumaczeniowych dla jednostek fatycznych i emotywnych	311
7.1 Metoda korpusowa	317
7.2 Metoda audiowizualna	367
7.3 Metoda semantyczno-pragmatyczna	382

ROZDZIAŁ VIII

8. Koncepcja artykułu hasłowego fatyzmów i ematywów	395
8.1 Tendencje w opisie artykułów hasłowych jednostek fatycznych i emotywnych w słownikach jedno- i dwujęzycznych	398
8.2 Potencjał dotychczasowych opracowań leksykograficznych w projektowaniu nowoczesnego artykułu hasłowego	402
8.3 Koncepcja makro- i mikrostruktury słownika jednostek fatycznych i emotywnych	411

ZAKOŃCZENIE	422

BIBLIOGRAFIA	426

STRESZCZENIE	445

SUMMARY	447

PEŽIOME	449

SHRNUTÍ	451

WPROWADZENIE

Współczesna leksykografia przekładowa, pomimo dynamicznego rozwoju i coraz szerszego wykorzystania narzędzi elektronicznych, wciąż nie uporała się z opisem jednostek wielowyrazowych. Tradycyjne podejście frazeologiczne doprowadziło do rejestracji w słownikach związków frazeologicznych i idiomów, jednak ze względu na ściśle określone cechy konstytutywne tychże jednostek wiele innych wielowyrazowców pozostawało poza obszarem badań językoznawczych. Tymczasem niezaprzeczalnym pozostaje fakt, iż język rzeczywisty, zwłaszcza jego odmiana mówiona, opiera się w dużej mierze na odtwarzaniu utrwalonych w użyciu zwrotów, konstrukcji i wyrażeń, które nie zawsze mieszczą się w ramach tradycyjnie pojmowanej frazeologii. Mimo postępu w badaniach leksykograficznych wciąż trudno uznać, że rejestracja jednostek charakterystycznych dla języka mówionego osiągnęła zadowalający poziom. Wynika to nie tylko z trudności metodologicznych w ich uchwyceniu, lecz także z przekonania, że odmiana mówiona stanowi mniej reprezentatywny wariant języka. Tego rodzaju jednostki, jak *świeżo malowane*; *wiadomość z ostatniej chwili*; *zakaz wjazdu* lub *czy o czymś nie zapomniałem*, stanowią istotną część zasobu leksykalnego języka, lecz przez długi czas pozostawały poza zakresem zainteresowań frazeologii opisowej.

Przez wskazane jednostki należy rozumieć stosowany już w językoznawstwie polskim, rosyjskim i czeskim termin *frazem*, pełniący funkcję hiperonimu dla tego typu struktur, lub pojęcie *reproduktu*, zaproponowane przez Wojciecha Chlebdę, w którym uwidacznia się kluczowa cecha tych jednostek – reprodukowalność. Problem braku ich rejestracji w opracowaniach leksykografii jedno- czy dwujęzycznej jest szczególnie widoczny w przypadku wyrażeń realizujących funkcje **fatyczne** (np. *wydaje mi się, że...*; *nawiasem mówiąc*; *chciałoby się rzec*) czy **emotywne** (np. *jak mi przykro*; *ale się cieszę*; *nie do wiary!*), które mimo swojej powszechności i wysokiego stopnia reprodukowalności rzadko stają się przedmiotem analiz leksykograficznych. Zagadnienie to nabiera szczególnego znaczenia w kontekście badań nad leksykografią przekładową, w której tradycyjnie dążono do prezentowania ekwiwalentów na poziomie jednostek systemowych, a nie komunikacyjnych. Ich adekwatny opis wymaga odejścia od tradycyjnej koncepcji słowa jako podstawowej jednostki słownikowej na rzecz szerzej rozumianych jednostek wielowyrazowych, które funkcjonują w języku jako całościowe wzorce komunikacyjne. Niniejsza praca stanowi próbę opisu i klasyfikacji wielowyrazowych jednostek fatycznych i emotywnych w języku polskim, rosyjskim i czeskim oraz ustalenia ich

ekwiwalentów przekładowych. Celem badań jest nie tylko rejestracja i analiza tych jednostek w kontekście jednojęzycznym, ale także refleksja nad możliwościami ich wykorzystania w nowoczesnej leksykografii przekładowej.

W ujęciu wspomnianego Wojciecha Chlebdy reprodukty stanowią podstawową kategorię opisu frazematyki, uwzględniającą jednostki funkcjonujące w języku jako niepodzielne całości znaczeniowo-komunikacyjne. Ich identyfikacja i opis wymagają jednak nie tylko świadomości teoretycznej, lecz także zastosowania odpowiednich narzędzi analitycznych, pozwalających uchwycić rzeczywiste użycie tych konstrukcji w komunikacji. W rozprawie wykorzystano zatem **metody korpusowe**, które umożliwiają weryfikację i egzemplifikację jednostek wielowyrazowych w autentycznych kontekstach. Podstawę materiałową stanowią **korpusy jednojęzyczne**, takie jak Narodowy Korpus Języka Polskiego, Narodowy Korpus Języka Rosyjskiego (*Национальный корпус русского языка*), Narodowy Korpus Języka Czeskiego (*Český národní korpus*) oraz **korpusy równoległe**, zwłaszcza InterCorp. Wskazane zasoby umożliwiają wyszukiwanie struktur wielowyrazowych z wykorzystaniem funkcji konkordancji, kolokatorów oraz analiz frekwencyjnych, a także obserwację ich występowania w różnorodnych kontekstach komunikacyjnych. Przyjęto jednocześnie założenie, że analiza korpusowa, choć niezwykle przydatna, nie stanowi wystarczającego źródła danych. Korpusy, oparte głównie o materiał pisany lub selektywnie rejestrowany, nie zawsze w pełni oddają dynamikę języka mówionego. W związku z tym podjęte badania uzupełniono o autorskie techniki wyszukiwania, opisu i ustalania ekwiwalentów przekładowych, pozwalające na precyzyjniejsze uchwycenie zjawiska fatyzmów i emotywów we wspomnianych językach. Przyjęta w rozprawie perspektywa jest kontynuacją myśli Andrzeja Bogusławskiego (1989), który wskazywał, że jednym z kluczowych zadań językoznawstwa pozostaje identyfikacja i opis wielowyrazowych jednostek języka jako nieodłącznych składników kompetencji komunikacyjnej użytkowników. Całość badań opiera się na dwóch ujęciach: **frazematycznym**, dotyczącym budowy i znaczenia jednostek, oraz **leksykograficzno-translatorycznym**, koncentrującym się na ich funkcjonowaniu w słownikach i możliwościach przekładu. Z przedstawionych założeń wynikają następujące **hipotezy badawcze**:

1. Jednostki wielowyrazowe o funkcji fatycznej i emotywnej w polszczyźnie, ruszczyźnie i czeszczyźnie nie zostały dotąd dostatecznie opisane w leksykografii jednojęzycznej ani przekładowej,

2. Korpusy językowe stanowią obecnie jedno z najskuteczniejszych narzędzi umożliwiających ich identyfikację i analizę, ponieważ rejestrują rzeczywiste użycia jednostek w konkretnym otoczeniu tekstowym,
3. Zastosowanie narzędzi korpusowych i autorskich metod analitycznych pozwala nie tylko na identyfikację jednostek, lecz również na ustalenie ich wariantów formalnych i semantycznych oraz wskazaniu adekwatnego przekładu.

Z przedstawionych hipotez wynikają również **pytania badawcze**, które wyznaczają kierunek dalszych analiz:

- a) Jakiego typu jednostki wielowyrazowe o funkcji fatycznej i emotywniej występują w języku polskim, rosyjskim i czeskim?
- b) W jaki sposób można dokonać ich skutecznej identyfikacji i klasyfikacji przy użyciu narzędzi korpusowych i pozakorpusowych?
- c) Jaki obiekt w sensie formalnym stanowi wielowyrazowiec i jak rozwiązać problem ich wariantowości w opisach leksykografii jedno- i wielojęzycznej?
- d) W jaki sposób ustanowić adekwatne ekwiwalenty dla jednostek emotywnych i fatycznych, które nierzadko charakteryzują się wysokim stopniem wieloznaczności?
- e) Jaką postać powinien przyjąć artykuł hasłowy w nowoczesnym słowniku przekładowym, który rejestruje jednostki fatyczne i ematywne?

Rozprawa ma charakter opisowo-analityczny i łączy refleksję teoretyczną z badaniami empirycznymi. Jej celem nadrzędnym jest przedstawienie metod wyodrębniania jednostek fatycznych i emotywnych w językach polskim, czeskim i rosyjskim, zaproponowanie sposobów ich przekładu oraz opracowania w formie nowoczesnego słownika przekładowego.

Struktura pracy została podporządkowana przyjętym celom i hipotezom badawczym, a jej układ odzwierciedla stopniowe przechodzenie od zagadnień teoretycznych do analizy empirycznej oraz rozwiązań aplikacyjnych. Praca składa się z ośmiu rozdziałów, z których każdy stanowi kolejny etap w procesie dochodzenia do wyników i wniosków końcowych.

Rozdział pierwszy poświęcony jest przedstawieniu genezy i rozwoju leksykografii polskiej, rosyjskiej i czeskiej. Zarysowano w nim kolejne etapy kształtowania się tradycji słownikarskich w trzech badanych językach, od wczesnych kompendiów i słowników normatywnych po współczesne opracowania elektroniczne. Szczegółowo przeanalizowano wybrane słowniki jednojęzyczne języka polskiego, rosyjskiego i czeskiego oraz słowniki przekładowe polsko-rosyjskie i polsko-czeskie, dokonując ich porównania pod względem sposobu prezentacji jednostek wielowyrazowych. Oceniono również zakres ich rejestracji i

sposoby definiowania, wskazując na różnice w podejściu do opisu materiału. Celem rozdziału jest ukazanie, że współczesne badania nad jednostkami wielowyrazowymi w leksykografii przekładowej są osadzone w długoletniej tradycji, lecz wymagają rewizji dotychczasowych metod opisu i rozszerzenia pola obserwacji o jednostki języka mówionego.

Rozdział drugi poświęcony jest reprodukowalności, stanowiącej kluczową kategorię w analizie jednostek fatycznych i emotywnych. Przedstawiono w nim rozwój refleksji nad jednostkami utrwalonymi – od tradycyjnej frazeologii po koncepcję frazematyki – ukazując przejście od opisu idiomów i związków frazeologicznych do szerszego ujęcia reproduktów językowych. Omówiono podstawowe założenia teoretyczne obu nurtów oraz ich reprezentatywne ujęcia w tradycji polskiej, rosyjskiej i czeskiej, między innymi w pracach Wiktora Winogradowa, Františka Čermáka, Andrzeja Marii Lewickiego i Wojciecha Chlebdy. Szczególną uwagę poświęcono frazematyce jako nowoczesnej koncepcji opisu jednostek o wysokim stopniu utrwalenia i reprodukowalności w języku.

Rozdział trzeci stanowi wprowadzenie do zagadnień lingwistyki korpusowej oraz prezentuje najważniejsze korpusy językowe wykorzystane w badaniach. Przedstawiono w nim główne założenia językoznawstwa korpusowego oraz omówiono strukturę i funkcjonalność następujących korpusów: Narodowy Korpus Języka Polskiego, Narodowy Korpus Języka Rosyjskiego, Narodowy Korpus Języka Czeskiego oraz MoncoPL. Opisano ich budowę, zakres, sposoby anotacji oraz dostępne narzędzia wyszukiwawcze, zwracając uwagę na możliwości wykorzystania zawartych w nich funkcji. Celem rozdziału jest ukazanie potencjału tych zasobów i wskazanie, w jaki sposób ich funkcjonalności mogą zostać wykorzystane w dalszych etapach pracy do wyszukiwania i analizy jednostek wielowyrazowych.

Rozdział czwarty poświęcony jest charakterystyce korpusów wielojęzycznych oraz omówieniu ich znaczenia w badaniach kontrastywnych i przekładowych. Przedstawiono w nim podział korpusów na porównywalne i równoległe oraz pokazano ich potencjał badawczy w analizie relacji międzyjęzycznych. Szczegółowo opisano strukturę i funkcjonalność korpusu InterCorp oraz Korpusu Równoległego Narodowego Korpusu Języka Rosyjskiego, uwzględniając ich budowę, system anotacji, zasady wyrównywania zdań i możliwości wyszukiwania ekwiwalentów tłumaczeniowych. Zwrócono uwagę na praktyczne aspekty wykorzystania korpusów w badaniach nad jednostkami wielowyrazowymi. Celem rozdziału jest ukazanie, w jaki sposób funkcjonalności korpusów wielojęzycznych mogą zostać zastosowane w dalszej części pracy do poszukiwania i weryfikacji ekwiwalentów tłumaczeniowych jednostek fatycznych i emotywnych.

Rozdział piąty służy omówieniu pojęcia re produktu, które stanowi podstawową kategorię opisu jednostek wielowyrzowych w rozprawie. Przedstawiono w nim źródła tej koncepcji oraz kierunki jej rozwoju w pracach Wojciecha Chlebdy, wskazując na znaczenie re produktów jako utrwalonych w języku formuł komunikacyjnych. Omówiono definicje i właściwości dwóch kluczowych podtypów re produktów – fatyzmów i emotywów – oraz opisano ich funkcje w języku mówionym. Fatyzmy zaprezentowano jako jednostki służące podtrzymywaniu kontaktu i regulowaniu przebiegu interakcji, natomiast emotywy jako wyrażenia o funkcji ekspresywnej, ujawniające emocjonalny stosunek nadawcy do wypowiedzi lub sytuacji komunikacyjnej. Istotnym elementem rozdziału jest przedstawienie autorskiej klasyfikacji tychże jednostek, opartej o kryteria semantyczno-pragmatyczne i funkcjonalne. Celem rozdziału jest uporządkowanie pojęć oraz wskazanie teoretycznych podstaw dla dalszej analizy materiału językowego i wyodrębniania ekwiwalentów tłumaczeniowych w językach polskim, rosyjskim i czeskim.

Rozdział szósty przedstawia zastosowane w pracy metody wyodrębniania re produktów fatycznych i emotywnych. W pierwszej kolejności omówiono metodę słownikowo-kolokacyjną, polegającą na łączeniu danych leksykograficznych i korpusowych w celu identyfikacji jednostek wielowyrzowych o utrwalonej strukturze. Następnie opisano wykorzystanie pomocników formalnych oraz metaoperatorów narracyjnych, które pełnią funkcję wskaźników odtwarzalności i pozwalają na nieautomatyczne wskazanie potencjalnych re produktów w tekście. Kolejne podrozdziały omawiają metodę synonimiczną i wyszukiwanie przyrostowe, umożliwiające rozszerzanie zbioru jednostek o warianty formalne i semantyczne. Rozdział zamyka opis metody tłumaczeniowej, opartej na analizie przekładów jako źródła poświadczeń re produktów. Całość rozdziału ma na celu ukazanie zróżnicowanych, komplementarnych sposobów pozyskiwania jednostek fatycznych i emotywnych, z uwzględnieniem ich funkcji pragmatycznych i potencjału translacyjnego.

Rozdział siódmy poświęcony jest metodom ustalania ekwiwalentów tłumaczeniowych dla jednostek fatycznych i emotywnych. Przedstawiono w nim różne sposoby pozyskiwania ekwiwalentów wykazujących najwyższy stopień zgodności z jednostką wyjściową – od metod korpusowych, poprzez analizę audiowizualną opartą na porównaniu wersji dubbingowych filmów animowanych, aż po podejście semantyczno-pragmatyczne. Wskazano również ograniczenia poszczególnych metod oraz zaproponowano ich komplementarne wykorzystanie w celu uzyskania najbardziej reprezentatywnych odpowiedników przekładowych.

Rozdział ósmy, stanowiący zwięźczenie pracy, zawiera propozycję nowoczesnego artykułu hasłowego dla słownika polsko-rosyjskiego i polsko-czeskiego, rejestrującego

jednostki wielowyrazowe o funkcji fatycznej i ematywnej. Rozdział ten łączy wnioski teoretyczne i empiryczne, ilustrując, w jaki sposób materiał badawczy może zostać opracowany w formie słownikowej zgodnie z zasadami współczesnej leksykografii przekładowej. Przedstawione rozwiązania odwołują się do najlepszych praktyk stosowanych w dotychczasowych opracowaniach leksykograficznych, wskazując jednocześnie możliwości ich adaptacji i rozwinięcia w kierunku rejestracji jednostek mówionych.

Zagadnienie jednostek wielowyrazowych o funkcji fatycznej i ematywnej sytuuje się na styku frazematyki, pragmatyki, translatoryki i leksykografii. Niniejsza rozprawa stanowi próbę zintegrowania tych perspektyw poprzez połączenie narzędzi korpusowych i refleksji nad praktyką słownikarską. W centrum uwagi znalazł się nie tylko sam opis jednostek, lecz także sposób ich reprezentacji w opracowaniach przekładowych, które wciąż w niewystarczającym stopniu odzwierciedlają rzeczywiste funkcjonowanie języka mówionego. Przeprowadzone analizy pozwalają spojrzeć na leksykografię przekładową jako na dynamicznie rozwijającą się dziedzinę, zdolną do łączenia tradycyjnych metod opisu z nowoczesnymi rozwiązaniami technologicznymi, jakie oferuje językoznawstwo korpusowe. Zaprezentowane w pracy wyniki badań mają również charakter aplikacyjny i mogą stanowić podstawę dalszych opracowań z zakresu słownictwa polsko-rosyjskiego i polsko-czeskiego, zwłaszcza w odniesieniu do rejestracji jednostek języka mówionego. Rozprawa wpisuje się tym samym w nurt współczesnych badań nad frazematyką i leksykografią przekładową, wskazując kierunek możliwego rozwoju tej dyscypliny w stronę modeli zorientowanych na rzeczywisty użycie językowy i komunikacyjną funkcjonalność jednostki. W ten sposób praca łączy wymiar teoretyczny z praktycznym, stanowiąc zarówno refleksję nad miejscem jednostek wielowyrazowych w słowniku przekładowym, jak i próbę stworzenia modelu ich opisu uwzględniającego funkcje komunikacyjne i translatoryczne.

ROZDZIAŁ I

Leksykografia: zakres, definicje, podziały

Leksykografia stanowi współcześnie jeden z najważniejszych działów językoznawstwa, wypełniając szereg społecznie istotnych funkcji, takich jak dydaktyczna, deskryptywna oraz normatywna, a zarazem wspiera próby rozwiązania złożonych problemów językoznawczych. W uproszczeniu przedmiotem pracy leksykograficznej jest słowo rozpatrywane we wszystkich swoich właściwościach, co sprawia, że opracowania słownikowe czerpią w szerokim zakresie z takich dziedzin, jak semantyka, składnia, fonologia, stylistyka czy pragmatyka. Potwierdza to M. Grochowski, według którego słowniki językowe opisują wyrażenia należące do jednego lub kilku języków naturalnych, analizując je z perspektywy różnych działów językoznawstwa (Grochowski 1982: 34–35). Z perspektywy funkcjonalnej procedurę leksykograficzną można sprowadzić syntetycznie do następującego opisu: „Lexikografie [...] podle jistého postupu zpracovává lexikální jednotky, slova a slovní spojení v jednotlivá hesla a sestavuje z nich slovníky různých druhů a typu” (Filipec 1957: 129).

Sam termin *leksykografia* pochodzi od greckich wyrazów *lexikos* oznaczającego *słownikowy* oraz *grapho* znaczącego *piszę*.

Pomimo długoletniej tradycji badań, definicja i zakres tego pojęcia do dziś nie są jednoznaczne i bywają przedmiotem dyskusji terminologicznych.

W literaturze przedmiotu przyjęto natomiast rozróżnienie pomiędzy leksykografią praktyczną a teoretyczną, które odzwierciedla dwa komplementarne aspekty tej dziedziny.

Pierwszy z nich należy traktować jako działalność, związaną z opracowywaniem konkretnych słowników i uwzględniającą zarówno dobór jednostek hasłowych, ustalanie ich struktury, jak i sposoby prezentacji informacji, a także tworzenie pierwotnych materiałów słownikowych, takich jak kartoteki językowe, korpusy tekstowe, glosariusze specjalistyczne czy elektroniczne bazy danych.

Z kolei leksykografia teoretyczna intensywnie rozwijana od drugiej połowy XX wieku, koncentruje się głównie na problematyce typologii słowników, opisu metajęzykowego, a także na zagadnieniach związanych z użytkownikiem słownika oraz ogólnych zasadach reprezentacji w nim leksyki.

Należy wspomnieć, iż dziedziną, z którą leksykografia jest najściślej powiązana, pozostaje leksykologia – to ona dostarcza podstaw do opisu, klasyfikacji i interpretacji

jednostek leksykalnych. W tradycyjnym ujęciu leksykologia skupia się na badaniu znaczeń, struktury i funkcjonowania wyrazów, natomiast leksykografia, traktowana jako dyscyplina stosowana, opracowuje dane na potrzeby praktycznego opisu słownikowego. W tym kontekście leksykografia przyjmuje często funkcję nadrzędną wobec leksykologii, czyli porządkuje i kodyfikuje informacje pochodzące z różnych działów językoznawstwa, i nadaje im przystępną formę dla użytkownika. Jednocześnie podkreśla się fakt, że choć leksykografia i leksykologia mają wspólny przedmiot badań, zakres opisu w przypadku tej pierwszej jest znacznie szerszy, zwłaszcza w odniesieniu do słowników jednojęzycznych:

„Má sice stejný předmět zkoumání jako lexikologie, t.j. lexikální zásobu, ale rozsah jejího popisu a výklad je zvláště u jednojazyčného slovníku výkladového širší – musí se zabývat např. i pravopisem slov a jejich výslovností, popř. i přízvukem, a kodifikuje i morfologickou vazebnou normu lexémů” (Filipec, Čermák 1986: 14).

Współczesne podejście leksykograficzne opiera się właśnie na ścisłej współpracy różnych dziedzin, od językoznawstwa i kulturoznawstwa po kognitywistykę czy technologię informacyjną, co nadaje współczesnemu słownikarstwu charakter silnie synkretyczny i interdyscyplinarny.

Jak zauważa B. Ju. Gorodieckij:

«Лексикография как научная дисциплина носит, несомненно, комплексный характер: её историко-филологический аспект проявляется в постоянном интересе к типологии словарей и их связи с культурой, её гносеологический аспект обусловлен сущностью словарей как способа организации и представления знаний, накопленных обществом, её семантико-лексикологический аспект связан с обобщением словарных данных в русле моделирования плана содержания языка. Но всё же определяющей чертой лексикографии является её прикладная направленность» (Городецкий 1983: 8–9).

W ujęciu Ju. D. Apresjana, charakterystyczną cechą całej współczesnej leksykografii jest synteza filologii i kultury w szerokim znaczeniu tego słowa, co podkreśla jej otwartość na konteksty humanistyczne i społeczne (Апресян 1993: 6). Pomimo faktu, iż praktyka tworzenia słowników sięga starożytnych glosariuszy i średniowiecznych kompilacji, dopiero w drugiej połowie XX wieku leksykografia wykształciła się jako samodzielna dyscyplina naukowa. Współcześnie zakres jej zastosowań uwzględnia, m.in. tłumaczenie, edukację językową, systemy informacyjno-wyszukiwawcze, badania kulturowe, etnograficzne, a także zagadnienia z pogranicza socjologii i psychologii. To właśnie interdyscyplinarność oraz użytkowy wymiar leksykografii nadają jej coraz większe znaczenie w warunkach intensywnych przemian technologicznych i cyfryzacji zasobów językowych.

Leksykografia, poza swoim interdyscyplinarnym i praktycznym charakterem, może być także postrzegana jako etap modelowania systemu językowego. Każdy twórca słownika dąży do określonego sposobu systematyzacji opisywanego materiału, konstruuje w jego ramach całą systematykę języka na poziomie zasobu leksykalnego. Można rzec, iż w pewnym sensie leksykograf rozwija przed użytkownikiem system języka, ponieważ «лингвистической реальности соответствует свёрнутый характер существования лексики как системы» (Денисов 1993: 29).

Słownik nie tylko odzwierciedla rzeczywistość językową, ale również ją porządkuje i interpretuje, nadając zgromadzonym jednostkom leksykalnym określoną strukturę i sens. W zależności od przyjętych założeń i funkcji można wyróżnić, m.in.

- **słowniki jednojęzyczne**, które objaśniają słownictwo w tym samym języku,
- **słowniki dwujęzyczne i wielojęzyczne** zawierające odpowiedniki przekładowe,
- **słowniki specjalistyczne** (np. terminologiczne, frazeologiczne, ortograficzne) oraz
- **ogólne**, uwzględniające leksykę codzienną i literacką.

Zróznicowanie słowników może dotyczyć również ich funkcji (deskryptywnej, normatywnej, dydaktycznej), sposobu uporządkowania haseł (alfabetyczny, tematyczny) czy grupy docelowych użytkowników.

Jedną z najbardziej charakterystycznych cech dzieł leksykograficznych jest ich wyraźnie ukształtowana makro- i mikrostruktura, w szczególności rozbudowany artykuł hasłowy, który we współczesnych słownikach przeważnie nie ogranicza się wyłącznie do definicji, lecz zawiera także dane gramatyczne, stylistyczne, etymologiczne, najczęstsze kolokacje, kwalifikatory oraz przykłady użycia w rzeczywistym kontekście (Żmigrodzki 2009: 63–79).

Wszystkie te cechy wynikają z coraz wyraźniejszego ukierunkowania leksykografii na odbiorcę i jego potrzeby informacyjne — to właśnie użytkownik staje się dziś centralnym punktem refleksji leksykograficznej. Jak zauważa R. Hartmann, to właśnie analiza oczekiwań i kompetencji użytkownika wyznacza kierunki rozwoju nowoczesnych metod opisu słownikowego (Hartmann 1989: 102–103).

Obok ujęć naukowych i funkcjonalnych podnoszona jest również kwestia twórczego charakteru leksykografii — tradycyjnie pojmowanej jako teoria i praktyka opracowywania słowników — zbliżającego ją do działalności artystycznej i znajduje wyraz w koncepcji *ars lexicographica*. Skonstruowanie słownika wymaga więc nie tylko wiedzy i doświadczenia, lecz

także intuicji językowej, wyobraźni i wyczucia stylu, co czyni tę dziedzinę pokrewną twórczości artystycznej. Jak zauważa W. W. Dubicinskij:

«Немаловажными качествами лексикографа мне представляются природные творческие способности, талант исследователя, чутьё языка в сочетании с желанием посвятить себя» (Дубичинский 2008: 44).

W podobnym duchu wypowiada się Tadeusz Piotrowski, pisząc:

„Chodzi o to, że w pracy leksykografa istnieją elementy niepoddające się opisowi scjentyistycznemu, nie oparte na logicznym rozumowaniu i wnioskowaniu, a podlegające intuicji, wyczuciu językowemu, przekonaniach płynących z doświadczenia” (Piotrowski 1994: 11).

S. Landau zaznacza, że tworzenie słowników nie wymaga błyskotliwości ani oryginalności myślenia, lecz opiera się na inteligencji, profesjonalizmie i poświęceniu w ciężkiej pracy (Landau 1989: 4).

Choć niektórzy autorzy podkreślają silnie techniczny charakter słownikarstwa, trudno pominąć fakt, że wiąże się ona również z decyzjami interpretacyjnymi, wymagającymi językowej intuicji i wyczucia.

Zwracał na to uwagę już L.W. Szczerba, podkreślając, iż działalność słownikowa, zwłaszcza oparta na analizie semantycznej, wymaga wyjątkowego daru postrzegania języka, w pewnym sensie pokrewnego talentowi pisarskiemu (Щерба 1958: 76). Istnieją naturalnie przeciwnicy tego poglądu, twierdzący, iż praca leksykografa w niewielkim stopniu ma charakter twórczy. Ich zdaniem opracowywanie słownika polega głównie na precyzyjnym stosowaniu ustalonego szablonu, a dobry opis leksykograficzny to taki, który jest spójny z innymi, także wcześniejszymi. Wyraz *sztuka* jako określenie działalności leksykografa występuje więc synonimicznie do pojęcia *rzemiosło*.

Można podsumować, iż leksykografia, stanowiąc samodzielny i interdyscyplinarny dział językoznawstwa, pełni istotną rolę nie tylko w porządkowaniu i systematyzowaniu wiedzy o języku, ale także o świecie, znajdującym w słowniku swoje odbicie. Współcześnie można zaobserwować nieustanny rozwój leksykografii, w której metody opisu i formy prezentacji materiału leksykalnego stale się udoskonalają i dostosowują do zmiennych realiów komunikacyjnych oraz potrzeb użytkowników. Słowniki – jako podstawowe narzędzie leksykografii – nie tylko odwzorowują zmiany języka, lecz wpływają na procesy kształtowania się świadomości językowej i kulturowej odbiorców. Choć słowniki tworzone są z uwzględnieniem bieżących realiów komunikacyjnych, zawsze odzwierciedlają stan języka

właściwy dla danego etapu jego rozwoju, a ich opracowanie wymaga nie tylko kompetencji naukowych i warsztatowych, lecz także wyczucia i precyzji, które wiążą się z intuicją i sztuką. W dalszej części pracy podjęta zostanie próba nakreślenia głównych etapów rozwoju leksykografii w Polsce, Czechach i Rosji – trzech tradycji, które odegrały istotną rolę nie tylko w kształtowaniu lokalnych świadomości językowych, lecz także w rozwoju europejskiej leksykografii jako całości. Opis poprzedzi poniższy zwięzły przegląd ogólnej historii leksykografii, ukazujący kluczowe momenty w jej rozwoju na gruncie światowym.

1.1 Zarys historii i rozwoju leksykografii

Początki leksykografii, rozumianej przede wszystkim jako działalność ukierunkowaną na systematyczne gromadzenie i objaśnianie jednostek językowych, sięgają czasów najdawniejszych cywilizacji piśmiennych, a jej ewolucja przebiega nieprzerwanie od ponad czterech tysięcy lat. Powstanie i rozwój dziedziny łączyć należy z potrzebą utrwalania, tłumaczenia i wyjaśniania semantyki słów zarówno wewnątrz jednej wspólnoty językowej, jak i w szeroko pojętych relacjach międzyjęzykowych. Przy czym jej zakres rozwinął się od glosariuszy o charakterze religijnym i edukacyjnym po współczesne, skomputeryzowane słowniki porządkujące wiedzę o języku i kulturze. Najwcześniejsze zarejestrowane formy działalności leksykograficznej wiążą się z cywilizacją sumeryjską, datowaną na około XXV wiek p.n.e., gdzie licznie powstawały pierwsze bilingwalne glosy i glosariusze, wykorzystywane zarówno w procesie kształcenia skrybów, jak i w praktykach administracyjnych oraz tłumaczeniowych, charakterystycznych dla zróżnicowanej językowo Mezopotamii. Analogiczne działania leksykograficzne podejmowano także w starożytnym Egipcie i Indiach, natomiast w Chinach już w III w. p.n.e. powstał encyklopedyczny leksykon tematyczny *Erya*, a kilka wieków później, w 121 r. n.e., filolog Xu Shen opracował pierwszy systematyczny słownik znaków chińskich, znany pod nazwą *Shuowen Jiezi* (Hanks 2013: 505). Z kolei w tradycji grecko-rzymskiej leksykografia przejawiała się głównie w postaci glos o charakterze marginalnym i interlinearnym, sporządzanych przez skrybów podczas przepisywania tekstów literackich, przy czym dopiero w okresie hellenistycznym i bizantyjskim zaczęto opracowywać obszerniejsze dzieła, takie jak np. normatywna i purystyczna *Eklogē* Frynicha z Bitynii. Również w Rzymie rozwijała się kultura leksykograficzna, czego świadectwem może być monumentalne dzieło *De Verborum Significatu* autorstwa Marka Weriusza Flakkusa, które na tle swojej epoki wyróżniało się nie

tylko dokumentacją zasobu leksykalnego, lecz także opisem aspektów kultury i wierzeń, stanowiąc tym samym cenne źródło wiedzy historycznej dla późniejszych pokoleń.

Warto również wspomnieć o rozwoju leksykografii w kulturze arabskiej, gdzie słowniki pełniły znaczącą rolę w procesie normatywizacji języka, interpretacji Koranu oraz wspieraniu twórczości poetyckiej, czego przykładem jest *Kitab al-Ayn* autorstwa Al-Chalila ibn Ahmada – jest to pierwszy znany systematyczny słownik języka arabskiego, w którym uporządkowano hasła według ich cech artykulacyjnych.

W kontekście średniowiecznej Europy leksykografia rozwinęła się od stosowania marginalnych glos do opracowywania rękopiśmiennych słowników łaciny z objaśnieniami w językach rodzimych, tworzonych głównie w klasztorach na potrzeby edukacyjne. Do najznamienitszych dzieł tego typu odnieść należy *Reichenauer Glossen* z VIII wieku (jeden z najstarszych glosariuszy łacińsko-romanojęzycznych) (Klein 1965: 182–183) oraz *Promptorium Parvulorum* z XV wieku (pierwszy obszerny słownik łacińsko-angielski wykorzystywany głównie w środowisku zakonnym) (Stein 1981: 317). Przełomowym momentem dla leksykografii okazało się wynalezienie druku w drugiej połowie XV wieku, które zrewolucjonizowało nie tylko sam proces tworzenia słowników, lecz także ich rozpowszechnianie, umożliwiając standaryzację, masowe powielanie oraz znacznie szerszy dostęp do wiedzy językoznawczej. W tym okresie drugie z wymienionych powyżej dzieł zyskało status jednego z pierwszych powszechnie dostępnych słowników i symbolicznie wyznaczyło początek nowej epoki w historii słownikarstwa. W okresie nowożytnym leksykografia uległa dalszemu rozwojowi i ugruntowaniu, czego wyrazem były przełomowe dzieła charakterystyczne dla poszczególnych kręgów językowych, takie jak *A Dictionary of the English Language* Samuela Johnsona (1755) w Anglii, *Encyclopédie* Diderota i d’Alemberta (1765) oraz *Grand Dictionnaire universel* Larousse’a w XIX-wiecznej Francji, *Deutsches Wörterbuch* braci Grimm (1852) w Niemczech czy słowniki Noaha Webstera w Stanach Zjednoczonych (Дубичинский 2008: 18).

W ciągu kolejnych dziesięcioleci, począwszy od przełomu XIX i XX wieku, leksykografia zaczęła stopniowo odchodzić od autorskich i indywidualnych form na rzecz działań instytucjonalnych, coraz silniej powiązanych z rozwojem nauk językoznawczych i systemem edukacji. Wraz z postępem metod badawczych i pojawieniem się narzędzi analitycznych opracowywano coraz liczniejsze słowniki specjalistyczne, dwujęzyczne czy dialektologiczne, a także rozbudowane dzieła o charakterze encyklopedycznym. Do najbardziej reprezentatywnych przykładów należą, m.in.: *Oxford English Dictionary*, zapoczątkowany pod

koniec XIX wieku i rozwijany przez kolejne dekady, oraz *Trésor de la langue française*, będący efektem obszernego opisu współczesnego języka francuskiego (Piotrowski 1994: 21).

Z kolei od drugiej połowy XX wieku coraz większe znaczenie zaczęły odgrywać technologie komputerowe, które początkowo pełniły funkcję pomocniczą, a z czasem stały się podstawowym narzędziem tworzenia, przechowywania i udostępniania słowników. Liczne prace elektroniczne, takie jak *Cambridge International Dictionary of English*, otworzyły nowe możliwości dostępu do treści oraz przeszukiwania ogromnych baz tekstowych według wybranych kryteriów. Współcześnie leksykografia opiera się przede wszystkim na lingwistyce korpusowej, która znajduje zastosowanie zarówno w opracowywaniu elektronicznych słowników jednojęzycznych, jak i w nowoczesnych słownikach dwu- oraz wielojęzycznych, umożliwiając precyzyjne odwzorowanie rzeczywistych kontekstów użycia słów w ich naturalnym otoczeniu językowym.

1.2 Rozwój leksykografii w Polsce

Historię rozwoju leksykografii na ziemiach polskich można podzielić na trzy zasadnicze etapy, przy czym należy zaznaczyć, że taki podział ma charakter umowny i nie stanowi powszechnie przyjętej klasyfikacji w badaniach naukowych. Wyróżnia się zatem:

- 1) okres najstarszy – do połowy XVI wieku,
- 2) etap od połowy XVI do końca XVIII wieku oraz
- 3) okres od początku XIX do połowy XX wieku (Żmigrodzki 2009: 126).

Pierwszy z wymienionych okresów związany był przede wszystkim z potrzebami nauczania religijnego i służył głównie celom dydaktycznym. Tworzone wówczas glosy i wokabularze stanowiły łacińsko-narodowe zestawienia słownictwa, które miały na celu ułatwienie nauki języka łacińskiego oraz zrozumienia tekstów sakralnych. Wśród zabytków piśmiennictwa leksykograficznego tego okresu można wymienić mamotrektory łacińsko-polskie (zbiór objaśnień do tekstów biblijnych), spośród których na szczególną uwagę zasługują *mamotrekt lubiński* i *mamotrekt kaliski*, przetrwałe do dzisiaj (Gruszczyński 2013: 207). Za cenne źródło leksyki specjalistycznej uznać należy również *Wokabularz trydencki* z około 1424 roku, rejestrujący blisko 500 haseł związanych z terminologią prawną, administracyjną oraz wojskową (Kaczyńska 2019: 67). Na przełomie XV i XVI wieku leksykografia przekładowa nadal dynamicznie się rozwijała, czego wyrazem było pojawienie się słowników trójjęzycznych, takich jak *Dictionarius variarum rerum* Jana Murmeliusza (1528) oraz *Dictionarium trium linguarum* Franciszka Mymera z tego samego roku. Tego rodzaju

opracowania, inspirowane wzorcami czeskimi i niemieckimi, miały przede wszystkim praktyczny charakter, a ich leksyka – rozszerzona o elementy potoczne, fachowe i religijne – odzwierciedlała dydaktyczno-przekładowe potrzeby i funkcje ówczesnej leksykografii.

W kolejnych stuleciach kontynuowano tradycję słowników przekładowych, rozwijając zarówno ich strukturę, jak i funkcje, czego przykładem może być *Lexicon latino-polonicum* Jana Mączyńskiego wydany w 1564 roku w Królewcu. Choć formalnie miał charakter łacińsko-polski, pełnił również funkcję przekładową i normatywną, a jego szczególną cechą – z perspektywy badań nad jednostkami wielowyrazowymi – było uwzględnienie przysłów, maksym i sentencji, nierzadko o starożytnym rodowodzie (Grześkowiak 2019: 192). Warto w tym miejscu wspomnieć również o monumentalnym dziele *Dictionarium decem linguarum* Ambrogia Calepina, którego dziesięciojęzyczna edycja, wydana w Lyonie w 1585 roku, objęła także język polski – obok m.in. angielskiego i węgierskiego. Spośród około 30 tysięcy haseł łacińskich aż 21 tysięcy miało polskie odpowiedniki, których autorstwo przypisuje się Piotrowi bądź Janowi Mączyńskiemu (Piotrowski 2001: 67). Obecność języka polskiego w tego typu słowniku świadczy o wzrastającej randze polszczyzny w europejskiej leksykografii przełomu XVI i XVII wieku. Wśród najważniejszych dzieł tego okresu wyróżnia się również *Thesaurus polono-latino-graecus* Grzegorza Knapiusza – pierwsze tak obszerne opracowanie (ok. 40 tys. haseł), w którym język polski po raz pierwszy został potraktowany jako wyjściowy, a łacinę i grekę podporządkowano jego strukturze. Słownik ten miał charakter normatywny – autor eliminował zapożyczenia i dążył do zachowania czystości języka, poprzez stosowanie neologizmów oraz słownictwa literackiego. Opracowanie to odznaczało się przejrzystą strukturą alfabetyczną, bogactwem frazeologizmów, synonimów oraz kwalifikatorów, wywierając tym samym znaczący wpływ na leksykografię polską aż do XIX wieku. Istotną rolę odegrał również trójjęzyczny słownik Michała Abrahama Trotza pod nazwą *Nowy dykcjonarz to jest Mownik polsko-francusko-niemiecki* z 1764 roku, który przyczynił się do utrwalenia nowatorskich zasad organizacji haseł słownikowych. Poprzez stosowanie kwalifikatorów i uwzględnianie informacji gramatycznych stał się ważnym etapem przejściowym między dawną leksykografią praktyczną a rozwijającą się leksykografią dydaktyczną i przekładową (Kuryłowicz 2021: 99–104).

Trzeci etap rozwoju leksykografii, przypadający na XIX i pierwszą połowę XX wieku, charakteryzuje się rosnącą dojrzałością warsztatu leksykograficznego, pojawieniem się słowników jednojęzycznych oraz próbą dokumentowania historii języka. Nie ulega wątpliwości fakt, iż najważniejszym dziełem tamtego okresu był *Słownik języka polskiego* autorstwa Samuela Bogumiła Lindego, liczący ponad 60 tys. haseł, w tym neologizmy i nazwy

własne. Linde stworzył słownik nie tylko opisowy, ale i porównawczy, zestawiając polskie wyrazy z odpowiednikami w językach słowiańskich i niemieckim, a jego dzieło o charakterze dokumentacyjnym i wychowawczym miało potwierdzać bogactwo polszczyzny w czasach zaborów. Zdecydowanie nowatorskim rozwiązaniem było opatrzenie haseł cytataми autorów od XVI wieku, co w ówczesnej leksykografii polskiej stanowiło zupełnie nowe podejście do dokumentowania znaczeń. W okresie tym powstawały również słowniki specjalistyczne oraz regionalne, a do najbardziej znanych można odnieść tzw. *Słownik wileński* wydany w 1861 roku. Praca ta łączyła tradycję słownika Lindego z nowszym podejściem rozwijanym w XX wieku, opierając się na dokumentowaniu i porządkowaniu źródeł oraz uwzględnieniu zróżnicowania stylistycznego i historycznego języka, co wywarło znaczny wpływ na rozwój późniejszej polskiej leksykografii jednojęzycznej (Żmigrodzki 2009: 140–141).

Pod koniec stulecia rozpoczęto natomiast prace nad tzw. *Słownikiem warszawskim* (1900–1927), który miał charakter deskryptywno-historyczny i stanowił kontynuację, a także rozwinięcie wcześniejszych tradycji słownikarskich. Przedstawiony szkic rozwoju leksykografii na ziemiach polskich ukazuje stopniowe przechodzenie od form elementarnych do coraz bardziej świadomych i złożonych koncepcji słownikowych. Etap powojenny, rozpoczynający się umownie po 1950 roku, otwiera nową fazę rozwoju leksykografii normatywnej i opisowej, a współczesne słowniki jednojęzyczne – zwłaszcza w kontekście uwzględnienia jednostek wielowyrazowych – będą przedmiotem analizy w niniejszej pracy.

1.3 Rozwój leksykografii w Rosji

Rozwój leksykografii rosyjskiej, podobnie jak polskiej, można podzielić na trzy etapy, różniące się formą słowników, ich funkcją oraz podstawami językowymi i ideowymi. Od średniowiecznych glosariuszy po współczesne opracowania, rosyjska tradycja słownikarska rozwijała się jako integralna część kultury piśmienniczej i naukowej, pozostając w ścisłym związku ze zmianami społecznymi, edukacyjnymi i politycznymi.

Pierwszy z tych etapów, określany jako przedśownikowy (XI–XVI w.), wiąże się ściśle z rozwojem piśmiennictwa cerkiewnosłowiańskiego. Już od XI wieku w przekładach ksiąg religijnych pojawiają się glosy i marginalia – dopiski objaśniające trudne słowa pochodzenia greckiego, hebrajskiego lub południowsłowiańskiego. Przykładem mogą być staroruskie glosy marginalne występujące w dziele *Изборник* z 1073 roku, zawierające objaśnienia niesłowiańskich i starobułgarskich wyrazów wprowadzone przez kopistę do tekstu w celu jego lepszego zrozumienia (Сороколетов 2001: 36). Do jednego z najważniejszych dzieł tego

okresu można odnieść glosariusz, dołączony do manuskryptu *Новгородская кормчая* z 1282 roku, który był sporządzony na potrzeby arcybiskupa Klimenta. W opracowaniu tym znalazły się 174 hasła objaśniające słownictwo greckie i hebrajskie z tekstów liturgicznych, a ich liczba w późniejszych redakcjach wzrosła do 344, co świadczy o stopniowym rozwoju tego typu opracowań (Тихомиров 1961: 8). Do pozostałych znaczących osiągnięć wczesnej leksykografii można odnieść *Новгородской словарь* z 1431 roku, stanowiący aneks do przekładu dzieła Jana Scholastyka, jednego z ojców Kościoła Wschodniego. W słowniku poszerzono zakres językowy i zwiększono liczbę jednostek z 61 do 200, a wykorzystanie serbskiego pierwowzoru wskazuje na żywe kontakty językowe z południowymi Słowianami.

Następny etap rozwoju, okreśłany umownie mianem wczesnej leksykografii (XVI–XVII w.), wiąże się z silnym rozwojem prac leksykograficznych w odpowiedzi na rosnące zapotrzebowanie na materiały edukacyjne i tłumaczeniowe, które były wykorzystywane głównie w szkolnictwie oraz praktyce religijnej. Upowszechnienie druku sprzyjało rozpowszechnianiu słowników i standaryzacji języka, wspierając główną funkcję leksykografii tego okresu, jaką było poznawanie i upowszechnianie literackiego języka pisanego, znacznie różniącego się od mowy potocznej, jak u wielu narodów średniowiecza. Należy wspomnieć, iż powstające wtedy słowniki charakteryzowały się synkretyzmem, tj. łączyły zarówno funkcje glosariuszy, przekładów, kompendiów jak i elementarzy, dążąc do wyjaśnienia trudnego słownictwa oraz kształtowania kompetencji językowych odbiorców. Na początku tego okresu dużą popularnością cieszyły się alfabetyczne glosariusze trudnych i obcych słów, zarówno pochodzenia słowiańskiego, jak i zapożyczonych z języków zachodnioeuropejskich, tzw. *азбуковники* (Гребенников 2017: 107). Ważny wkład w rozwój tego typu słowników wniósł Maksim Grek, który w swojej pracy *Толкование имен греческих и еврейских и римских по алфавиту* porządkował i objaśniał imiona pochodzenia biblijnego, przyczyniając się do utrwalenia alfabetycznego modelu opracowania leksykograficznego (Ковтун 1975: 5). Za przełomowy moment w dziejach rosyjskiej leksykografii należy uznać wydanie *Лексис* z 1596 roku autorstwa Lawrentija Zizanija Tustanowskiego – pierwszego drukowanego słownika na ziemiach wschodniosłowiańskich. Dzieło to, liczące 1061 haseł, służyło głównie wyjaśnianiu trudnych słów cerkiewnosłowiańskich przy użyciu prostego języka rosyjskiego, tłumacząc także wyrazy greckie i łacińskie (Дубичинский 2008: 19).

W podobnym duchu ukraiński leksykograf Pamvo Berinda skonstruował w 1627 roku *Лексикон славяноросский и имен толкование* – znacznie obszerniejszy, bo liczący około 7000 haseł, słownik wyjaśniający trudne i obce słowa występujące w tekstach religijnych i literackich. Z czasem obok glosariuszy i słowników o charakterze dydaktycznym zaczęły

pojawiać się pierwsze słowniki specjalistyczne. Rozwijało się słownictwo onomastyczne, czego przykładem może być *Речь жидовского языка* – zestawienie imion własnych, głównie hebrajskich, zaczerpniętych z Biblii. Równolegle kształtowała się leksykografia symboliczna, reprezentowana przez tzw. *приточки*, czyli słowniki wyjaśniające metaforyczne i alegoryczne znaczenia słów w kontekście tradycji egzegetycznej. Jednym z najważniejszych dzieł tego nurtu jest *Толк о неразумных словесех псалтырных*, zawierający objaśnienia słów o ukrytym, metaforycznym sensie, charakterystycznym dla duchowej tradycji Rusi.

W drugiej połowie XVII wieku pojawiły się przekładowe słowniki słowiańsko-ruskie, wyjaśniające trudne fragmenty tekstów religijnych, jak np. *Толкование неудобь познаваемом в писаных речем*, który koncentrował się na metaforach i figurach retorycznych. Równolegle rozwijały się pierwsze rozmówki – najwcześniejsze formy słowników przekładowych – m.in. *Речь тонкословия греческого* i *Толкование языка половецкого*, służące tłumaczeniu z języków obcych na ruski. Rezultatem tych tendencji było dzieło autorstwa Fiodora Polikarpowa Orłowa pod nazwą *Лексикон треязычный* z 1704 roku, który stanowił trójjęzyczny słownik rosyjsko-łacińsko-niemiecki, łączący funkcje dydaktyczne z dążeniem do standaryzacji języka (Козырев, Черняк 2000: 19). Niniejsze dzieło, jako jedno z najważniejszych w swoim czasie, można uznać za umowny początek trzeciego etapu rozwoju słownictwa rosyjskiego, który określamy mianem dojrzałej leksykografii (XVIII–XX w.).

Ostatni omawiany tu etap rozwoju leksykografii wiąże się z ugruntowaniem ogólnonarodowego języka literackiego oraz rosnącą potrzebą jego opisu i normatywizacji. Słowniki tego okresu służą już nie tylko dokumentowaniu zasobu leksykalnego, lecz także jego kodyfikacji oraz kształtowaniu świadomości językowej społeczeństwa. Rozwój słownictwa w tym okresie, zapoczątkowany przez reformy cara Piotra I i europeizację rosyjskiej kultury, był związany z intensyfikacją kontaktów międzynarodowych oraz rozwojem druku, czego skutkiem było upowszechnienie słowników dwujęzycznych, terminologicznych, a także obcojęzycznych. Przełomowym dziełem był *Словарь Академии Российской* (1789–1794), pierwszy normatywny słownik języka rosyjskiego, który – zawierając ponad 43 tysiące haseł, bogaty materiał frazeologiczny oraz cytaty – wyznaczył nowy kierunek w leksykografii, stając się prekursorem systematycznego, wieloaspektowego opisu języka. Z kolei w XIX wieku rosyjska leksykografia rozwijała się dynamicznie, czego dowodem są różne typy słowników – od etymologicznych po specjalistyczne. Przykładem może być *Словарь церковнославянского и русского языка* z 1874 roku, który jako pierwszy uwzględnia zarówno język żywy, jak i historyzmy, a także odzwierciedla nową tendencję polegającą na tworzeniu kartotek i dokumentacji cytatów, stającą się wówczas dominującą metodą pracy słownikowej.

Szczytowym osiągnięciem XIX-wiecznej leksykografii był czterotomowy *Толковый словарь живого великорусского языка* Władimira Dala, który opierał się o autentyczny język mówiony i liczył ponad 200 tysięcy haseł oraz tysiące przysłów i frazeologizmów. Swoją zawartością i układem przypominał encyklopedię rosyjskiego życia ludowego, odzwierciedlającą leksykę różnych warstw społecznych, środowisk zawodowych i regionów. W podobnym nurcie rozwijała się również leksykografia dialektalna i historyczna, czego wyrazem stało się monumentalne, wydawane pośmiertnie od 1893 roku dzieło Izmajła Srezniewskiego *Материалы для словаря древнерусского языка по письменным памятникам*, stanowiące fundamentalny wkład w badania nad dawnymi zasobami leksyki. Zaprezentowany szkic rozwoju rosyjskiej leksykografii do końca XIX wieku ukazuje wyłącznie najważniejsze etapy kształtowania się tradycji słownikarskiej, które przygotowały grunt pod kolejny okres rozpoczęty po 1917 roku. Następny etap, obejmujący wiek XX i początek XXI, nie będzie tu szczegółowo omawiany, lecz zostanie uwzględniony w analizie współczesnych słowników jednojęzycznych, rejestrujących jednostki wielowyrazowe.

1.4 Rozwój leksykografii w Czechach

Za umowny początek czeskiego słownictwa można uznać XIII wiek, kiedy to działalność leksykograficzna była ściśle związana z praktyczną potrzebą rozumienia tekstów pisanych w językach obcych, przede wszystkim po łacinie. Można stwierdzić, iż był to etap przygotowawczy prac słownikowych, polegający na zapisywaniu glos przekładowych i interpretacyjnych w tekstach obcojęzycznych – podobnie jak miało to miejsce w leksykografii polskiej czy rosyjskiej. Świadectwem tej działalności są liczne rękopisy czy wczesne druki z czeskimi glosami, do których należą, m.in. glosy w słowniku *Mater verborum* z XIII wieku (Kapusta 2008: 17). Część autorów skupiała się na wybranych dziedzinach tematycznych, tworząc uporządkowane zestawienia słownictwa, które stały się wzorem dla późniejszych słowniczków terminologicznych. Szczególne znaczenie w tym okresie miały zbiory glos z łacińskich zielników, tzw. *rostlináře*, spośród których do najstarszych należy *rostlinář Olomoucký* z początku XIV wieku. Rozwój średniowiecznego czeskiego słownictwa był stopniowy i zawierał m.in. *mamotreky* – zbiory trudniejszych słów biblijnych, porządkowane lokalnie (*Mamotrekt Kapitulní*) lub alfabetycznie (*Mamotrekt Klementinský*) (Hladká 2005: 140–159). Obok nich funkcjonowały również proste łacińsko-czeskie zestawienia przekładowe, występujące pod nazwami *vokabula* lub *vokabuláře*, a także różnorodne podręczniki do nauki łaciny, takie jak *sekvenciáře*, *hymnáře*, *poenitenciáře*, *roháče* czy *synonymáře*. Pomimo swojej wartości edukacyjnej i praktycznej, opracowania te nie spełniały

jeszcze typowych kryteriów słownika z uwagi na brak jednolitego korpusu językowego oraz makro – i mikrostruktury haseł. Z czasem zaczęto jednak tworzyć bardziej uporządkowane formy, czego wyrazem były chociażby alfabetyczne zbiory słów (*alfabetáře*) i zbiory tematyczne (*nomenklátory*). W ramach tych ostatnich szczególne znaczenie zyskały dzieła Bartłomieja z Chlumca, znanego jako Klaret, który w drugiej połowie XIV wieku opracował trzy łacińsko-czeskie słowniki: *Vokabulár*, *Bohemár* i *Glosár*. Zostały opracowane z myślą o studentach Uniwersytetu Karola i przybrały formę rymowanych *nomenklatorów*, przeznaczonych do nauki pamięciowej (Voleková 2013: 13). Jednocześnie stanowiły pierwszą próbę stworzenia spójnej czeskiej terminologii specjalistycznej, szczególnie w zakresie gramatyki. Mimo że wiele neologizmów Klareta się nie przyjęło, jego dzieła przez kolejne stulecia stanowiły podstawę późniejszej leksykografii czeskiej i były wykorzystywane aż do czasów odrodzenia narodowego. Koniec XV wieku przynosi koniec rękopiśmiennych słowników łacińsko-czeskich, tworzonych prozą lub wierszem, w których tylko nieliczne podejmowały próbę tworzenia czeskiej terminologii, a bardziej systemowe rozwiązania zostają wprowadzone w wiekach późniejszych.

W kolejnym okresie, tj. w XVI wieku, czeskie słownictwo przeżywa znaczący rozkwit, na co wpływ miały zarówno wynalezienie druku i laicyzacja oświaty, jak i idee reformacyjne oraz rosnące zainteresowanie językami narodowymi. W środowisku czeskim pojawiają się próby całościowego opisu zasobu leksykalnego języka ojczystego, a słowniki, dotąd pełniące głównie funkcję podręczników do tłumaczeń, stopniowo przekształcają się w narzędzia językoznawcze, co odzwierciedla się m.in. w tworzeniu pierwszych słowników etymologicznych. Do najważniejszych opracowań tego okresu odnieść należy dzieło pod nazwą *Lactifer* Jana Vodňanského z 1511 roku, będący pierwszym drukowanym słownikiem łacińsko-czeskim, który łączył porządek rzeczowy i alfabetyczny (Fidlerová i in., b.r.: 56) [dostęp 19.10.2025]. Warto również wymienić *česko-latinský slovník* Tomáša Rešela z 1562 roku, opracowany jako pomoc szkolna na podstawie dzieła Dasypodiusa, lecz wzbogacony o liczne rozwiązania autorskie. Jednakże najważniejszym osiągnięciem tego okresu były dzieła Daniela Adama z Veleslavína, który jako pierwszy nadał językowi czeskiemu nadrzędną pozycję w słowniku i konsekwentnie dążył do jego całościowego opisu. Jego prace *Dictionarium linguae latinae*, *Nomenclator tribus linguis*, *Nomenclator quadrilinguis* i *Sylva quadrilinguis* (Vajdlová 2019: 1–22), oparte na zagranicznych wzorcach, wprowadziły do czeskiego słownictwa nowoczesne rozwiązania metodologiczne, takie jak porządek alfabetyczny według czeskiego alfabetu, gniazdowanie haseł oraz rozbudowane ujęcie synonimii i frazeologii. To właśnie prace tego autora należy uznać za właściwy początek

czeskiej leksykografii, a ostatnie z wymienionych dzieł zachowało swoją wartość aż do czasów odrodzenia narodowego, będąc wielokrotnie przepisywane i wykorzystywane jako źródło.

Następny ważny etap rozwoju czeskiej leksykografii, przypadający na XVII wiek, to przede wszystkim okres działalności Jana Amosa Komenského, który kontynuował humanistyczną tradycję i dążył do całościowego ujęcia czeskiego zasobu językowego. Jego największy projekt *Thesaurus linguae Bohemicae* nie został wprawdzie zrealizowany, ale sam autor pozostawił po sobie ważne dzieła leksykograficzne, takie jak *Janua linguarum reserata* (1631) i *Orbis sensualium pictus* (1658), będące wielojęzycznymi podręcznikami o charakterze słowników rzeczowych. W jego dorobku znajduje się także zbiór przysłów *Moudrost starých Čechů za zrcadlo vystavená potomkům*, wydany później wraz z pracą *Didaktika česká*. Okres po bitwie pod Białą Górą przyniósł nasilony ucisk narodowościowy, ograniczenie funkcji języka czeskiego oraz jego systematyczne wypieranie ze sfery kultury i życia publicznego. W odpowiedzi na te procesy pojawiły się tendencje purystyczne, dążące do oczyszczenia języka z obcych, zwłaszcza niemieckich wpływów, co wpłynęło również na zmianę koncepcji leksykograficznych. Na tym tle wyróżnia się działalność Václava Jana Rosy, którego czterotomowy *Thesaurus linguae Bohemicae* stanowił istotny krok w stronę jednojęzycznego, objaśniającego słownika narodowego. Choć dzieło Rosy pozostało w rękopisie i nie całkiem uniknęło wpływów purystycznych, jego czesko-łacińsko-niemiecka struktura, licząca 30–40 tysięcy haseł uporządkowanych według gniazd etymologicznych, wprowadzała nowatorskie podejście do opisu semantyki, słowotwórstwa i – po raz pierwszy – także gramatyki. Łacina i niemiecki pełniły w tym słowniku funkcję pomocniczą, natomiast w centrum uwagi znajdowały się znaczenia jednostek czeskich, związki etymologiczne oraz zagadnienia strukturalne.

Jeden z najważniejszych okresów czeskiej historii, czyli odrodzenie narodowe, przyniosło intensywny rozwój czeskiej leksykografii na przełomie XVIII i XIX wieku i było napędzane dążeniem do przekształcenia czeszczyzny w nowoczesny język kultury i nauki. Rosnące zainteresowanie językiem doprowadziło do powstania licznych opracowań leksykograficznych o zróżnicowanej wartości, ponieważ obok cennych prac naukowych pojawiały się również te przepełnione sztucznymi neologizmami i nadmiernym puryzmem. Do ważniejszych prac tego czasu należą trzytomowy *Německo-latinsko-český slovník* Kašpara Zachariáša Vusína, silnie neologiczny *Slovář český* Jana Františka Ryvoli oraz słowniki Karla Ignácego Tháma. Szczególne znaczenie mają także opracowania Františka Jana Tomsy, zwłaszcza *Vollständiges Wörterbuch der böhmischen, deutschen und lateinischen Sprache* (1791), a także dzieła Josefa Dobrovského, sprzeciwiającego się nieuzasadnionemu neologizowaniu i jako pierwszy oparł słownik na rzeczywistym uzusie językowym (*Deutsch-böhmisches Wörterbuch*, 1802).

Zwieńczeniem leksykografii czeskiej okresu odrodzenia narodowego i fundamentem nowoczesnej czeszczyzny literackiej był pięciotomowy *Slovník česko-německý* Josefa Jungmanna (1834–1839), który łączył funkcję przekładową z opisem semantycznym i uwzględniał cały dostępny zasób leksykalny (Červená i in. 1982: 233–247). Dzieło to kontynuował František Ladislav Čelakovský, planując słownik oparty na rdzeniach, jednak ostatecznie opracował jedynie niewielki dodatek (1851), a zgromadzony materiał wykorzystał później František Štěpán Kott w dziesięciotomowym słowniku czesko-niemieckim (czes. *Česko-německý slovník*) (1878–1906). Prace Jungmanna i jego następców zamykają etap leksykografii odrodzenia narodowego i tworzą podstawy dalszego rozwoju czeskiego słownictwa naukowego i praktycznego. Na przełomie XIX i XX wieku czeska leksykografia wkroczyła w nowy etap rozwoju, w którym zaczęły się wyraźnie kształtować dwa główne nurty: leksykografia jednojęzyczna oraz specjalistyczna. Jej symbolicznym początkiem stał się *Průruční slovník jazyka českého* (1935–1957) – pierwszy ukończony jednojęzyczny słownik objaśniający języka czeskiego, który zapoczątkował standard opisu semantycznego w ramach spójnej mikrostruktury i stworzył podstawy dalszych badań nad czeszczyzną. Przedstawiony tu szkic dziejów czeskiego słownictwa miał na celu jedynie zarysowanie najważniejszych etapów jego rozwoju i wprowadzenie w kontekst historyczny, natomiast w dalszej części pracy uwaga zostanie skierowana wyłącznie na wybrane współczesne słowniki, istotne z punktu widzenia analizy jednostek wielowyrazowych.

1.5 Leksykografia jednojęzyczna wobec wielowyrazowych jednostek języka mówionego

Leksykografia jednojęzyczna, traktowana jako dziedzina zajmująca się opracowywaniem słowników danego języka, odgrywa istotną rolę nie tylko w dokumentowaniu i opisie systemu językowego, lecz także w jego współkształtowaniu oraz budowaniu świadomości językowej odbiorców. W tradycji słownikarskiej przyjęło się, że opracowania jednojęzyczne mają spełniać **dwie zasadnicze funkcje**, tj. nie tylko odzwierciedlać aktualny stan języka, lecz również wyznaczać normy jego poprawnego użycia. Przez wiele epok rozwoju leksykografii przeważała funkcja normatywna, przejawiająca się w selekcji, porządkowaniu i wartościowaniu jednostek leksykalnych (Žmigrodzki 2009: 112). W ostatnich latach można zaobserwować tendencję do odchodzenia od preskrytywizmu oraz eliminowania niepożądanych jednostek na rzecz empirycznego opisu, opartego na danych korpusowych i rzeczywistym uzusie językowym. Na tle tych przemian szczególne znaczenie zyskują słowniki

ogólne, które – jako powszechnie wykorzystywane opracowania przeznaczone dla szerokiego grona użytkowników – wywierają istotny wpływ na kształtowanie potocznego wyobrażenia o języku i jego normie (Дубичинский 2008: 29). To właśnie one dostarczają podstawowych informacji o znaczeniu, użyciu i poprawności jednostek leksykalnych, odgrywając tym samym kluczową rolę w upowszechnianiu określonego modelu kompetencji językowej oraz pośrednicząc między opisem systemowym a praktyką komunikacyjną. Uwzględniając fakt, że słowniki ogólne są najczęściej pierwszym źródłem, do którego użytkownicy języka sięgają w poszukiwaniu informacji o konkretnej jednostce, na autorach tychże opracowań spoczywa szczególna odpowiedzialność, powiązana z kształtowaniem wyobrażenia odbiorców o normie i sposobie funkcjonowania języka.

Celem podrozdziału nie jest jednak szczegółowe omówienie podstawowych założeń leksykografii jednojęzycznej, metod jej opisu czy typologii słowników jednojęzycznych. Poniższe rozważania mają charakter wprowadzający do dalszej analizy i służą przede wszystkim zwróceniu uwagi na **problem marginalizacji wielowyrazowych jednostek języka mówionego** w leksykografii jednojęzycznej polskiej, rosyjskiej i czeskiej. Zjawisko to dotyczy zwłaszcza najważniejszych i najbardziej wpływowych słowników jednojęzycznych opracowanych w środowiskach polskim, czeskim i rosyjskim. Pomimo szerokiego zakresu oraz znaczenia dla opisu i rozwoju leksykografii rzadko rejestrują odtwarzalne i wielowyrazowe jednostki języka mówionego oraz potocznego lub ujmują je w sposób bardzo ograniczony, choć wśród nich można wskazać także dzieła, które częściowo przełamują tę tendencję. Jednym z podstawowych założeń tradycyjnej leksykografii jednojęzycznej jest traktowanie **wyrazu** jako elementarnej jednostki artykułu hasłowego. Ujęcie to prowadzi do zjawiska określanego mianem **wyrazocentryzmu** (Chlebda 2023: 860–861), czyli przekonania, że język składa się przede wszystkim z wyrazów, a nie z większych, odtwarzalnych segmentów, takich jak reproduktory czy frazemy, z których to właśnie termin *frazem* jest powszechnie stosowany w językoznawstwie rosyjskim i czeskim. Chociaż już w połowie XX wieku pojawiały się krytyczne głosy wobec tak uproszczonego postrzegania języka, w praktyce leksykograficznej dominowało podejście, w którym jednostki wielowyrazowe były ujmowane wybiórczo, głównie w postaci idiomów, częstych kolokacji, przysłów, frazeologizmów.

Pomijanie takich jednostek wynikało i nadal wynika z kilku zasadniczych przyczyn, o których będzie mowa w rozdziale szóstym, poświęconym metodom wyodrębniania reproduktów.

Po pierwsze, trudność sprawia precyzyjne wyznaczenie granic jednostki wielowyrazowej, gdyż często nie sposób jednoznacznie określić, gdzie zaczyna się i kończy

ustabilizowany ciąg językowy. W przeciwieństwie do jednostek jednowyrazowych, których granice są z góry określone przez ich formę, struktury wielowyrazowe cechują się zmiennością i elastycznością, co utrudnia ich jednoznaczne ujęcie.

Po drugie, jednostki charakterystyczne dla języka mówionego mają często charakter potoczny, kolokwialny i silnie nacechowany emocjonalnie, przez co bywają postrzegane jako nieprzystające do języka ogólnego i tym samym niepretendujące do uwzględnienia w słowniku, który z definicji ma przedstawiać język ogólny.

Po trzecie, trudność stanowi znaczna wieloznaczność tego typu jednostek, wymagająca dodatkowej analizy, interpretacji i precyzyjnego opisu, co bywa trudne do uchwycenia w ramach tradycyjnych, papierowych opracowań leksykograficznych. Powiązane są z tym właśnie ograniczenia objętościowe i konieczność selekcji materiału, które w przypadku słowników drukowanych zmuszały leksykografów do zawężania zakresu haseł i wyboru tych uznawanych za najbardziej reprezentatywne z perspektywy języka ogólnego.

Wreszcie po czwarte, u podstaw wielu decyzji leksykograficznych leży dążenie do uporządkowania języka zgodnie z jego „czystą” i ustandaryzowaną postacią, co skutkuje eliminowaniem elementów uznawanych za nieregularne, chaotyczne czy odbiegające od normy. Tymczasem wielowyrazowe jednostki języka mówionego są często powiązane z operatorami metatekstowymi i to właśnie dzięki nim mogą realizować jedno ze swoich znaczeń.

Tego rodzaju praktyki pociągają za sobą istotne skutki zarówno poznawcze, jak i edukacyjne, ponieważ użytkownik słownika, zwłaszcza osoba ucząca się języka lub traktująca słownik jako podstawowe źródło wiedzy o jego strukturze i użyciu, może odnieść wrażenie, że język to przede wszystkim zbiór wyrażeń oddzielonych i samodzielnych wyrazów. W takiej perspektywie eliminowane są te elementy, odgrywających rolę szczególnie istotną w rzeczywistej komunikacji, czyli gotowe ciągi językowe, pełniące określone funkcje w relacjach społecznych i w interakcjach werbalnych. Pomijanie tego rodzaju jednostek w słownikach jednojęzycznych nie tylko utrudnia ich rozpoznawanie i rozumienie, lecz także wpływa na sposób postrzegania języka jako systemu, którego struktura ogranicza się do wyrazów i ich znaczeń. Warto jednak zauważyć, że nie wszystkie słowniki jednojęzyczne w pełni realizują ten wyrazocentryczny i normatywny model opisu, ponieważ w części nowszych prac leksykograficznych pojawiają się zmiany w podejściu do prezentacji i doboru materiału. Coraz częściej dostrzega się potrzebę rejestrowania jednostek wielowyrazowych, a także form właściwych dla żywego języka mówionego, w tym również konstrukcji nietypowych, wcześniej nie mieszczących się w ramach tradycyjnych opracowań słownikowych. Chociaż tendencje te nie mają jeszcze charakteru dominującego, można zaobserwować stopniową

ewolucję podejścia do opisu języka oraz rosnącą rolę danych empirycznych i obserwacji uzusu. W dalszej części przeanalizowane zostaną wybrane polskie, rosyjskie i czeskie słowniki jednojęzyczne pod kątem obecności lub braku jednostek fatycznych i emotywnych. Refleksja ta stanie się punktem wyjścia do porównania z praktyką leksykografii przekładowej, w której, mimo teoretycznie większej otwartości na różnorodność jednostek, także często dochodzi do marginalizacji interesujących nas reproduktów.

1.6 Rejestracja jednostek fatycznych i emotywnych we współczesnych polskich słownikach jednojęzycznych

Refleksję nad obecnością jednostek wielowyrazowych (zwłaszcza fatycznych i emotywnych) we wybranych współczesnych słownikach jednojęzycznych poprzedzi w danym miejscu zarys najważniejszych teorii leksykograficznych, które miały bezpośredni lub pośredni wpływ na polską myśl słownikarską od drugiej połowy XX wieku po współczesność. Zmieniające się podejścia do struktury, funkcji i celu słownika znalazły odzwierciedlenie w pracach najważniejszych polskich leksykografów, co znacząco wpłynęło na wybór, opis i sposób prezentacji materiału leksykalnego. Przedstawione zostaną tu wybrane, reprezentatywne podejścia: od klasycznego modelu normatywno-deskryptywnego, przez orientację pragmatyczną i tekstocentryczną, po bardziej współczesne koncepcje uwzględniające, m.in.: język potoczny i ekspresywny.

Zaprezentowane zarysy koncepcji stanowić będą punkt odniesienia dla analizy wybranych jednojęzycznych słowników języka polskiego opracowanych pod koniec XX i na początku XXI wieku, które zostaną zanalizowane pod kątem obecności lub nieobecności w nich jednostek wielowyrazowych, sposobu ich klasyfikacji, definicji oraz ich miejsca w strukturze artykułu hasłowego. Dobór ten uwzględnia zarówno słowniki ogólne, jak i te, które ze względu na zakres tematyczny lub swoją funkcję, mogą uwzględniać jednostki wielowyrazowe oraz materiał z języka mówionego. Zasadniczym celem będzie ustalenie, w jakim zakresie współczesna leksykografia jednojęzyczna uwzględnia reprodukty i utrwalone ciągi wyrazowe, w szczególności te o funkcji fatycznej i ematywnej.

Jedną z kluczowych postaci polskiej leksykografii XX wieku pozostaje Witold Doroszewski, redaktor wielotomowego *Słownika języka polskiego* (1958–1969) i twórca modelu, który na długi czas ukształtował zarówno praktykę leksykograficzną, jak i społeczne wyobrażenia o formie słownika ogólnego. Jego koncepcja, rozwijana m.in. w pracach z lat 1954 i 1970, opierała się głównie na założeniach behawiorystycznych, scjencycznych i

racjonalistycznych. Fundamentem teorii Doroszewskiego było przekonanie, że definicja słownikowa powinna perswazyjnie oddziaływać na zachowania językowe użytkownika poprzez wskazanie prawidłowych sposobów użycia wyrażen w konkretnych sytuacjach komunikacyjnych. W jego opinii słownik miał pełnić więc nie tylko funkcję informacyjną, ale przede wszystkim dydaktyczno-wychowawczą. Zgromadzony materiał językowy był selekcionowany według kryteriów staranności i poprawności, czego wyrazem było marginalizowanie słownictwa z niższych rejestrów stylistycznych oraz ograniczenie form potocznych, ekspresywnych czy środowiskowych. Charakterystyczne dla danej teorii było również dążenie do encyklopedyzmu – definiowania nie tylko znaczeń słów, lecz także odpowiadających im desygnatów i realiów pozajęzykowych. Jednostki wielowyrazowe nie były w jego słowniku ujmowane jako samodzielne hasła, lecz pojawiały się jedynie sporadycznie – głównie jako cytaty, przykłady użycia lub elementy bardziej rozbudowanych konstrukcji frazeologicznych. Taki model, często określany mianem „typu Doroszewskiego”, przez wiele lat wyznaczał standardy, stając się punktem odniesienia dla kolejnych słowników jednojęzycznych języka polskiego.

Należy w tym miejscu wspomnieć również o Stanisławie Skorupce, językoznawcy i leksykografie, autorze jednego z najważniejszych dzieł polskiej frazeologii XX wieku. Wydany w 1967 roku *Słownik frazeologiczny języka polskiego* koncentruje się, co prawda, na dokumentacji związków frazeologicznych i nie podejmuje zagadnienia opisu języka ogólnego, jednak w kontekście analizowanej problematyki jego rola zasługuje na szczególną uwagę. Z punktu widzenia badań nad wielowyrazowymi reproduktami, zwłaszcza o charakterze fatycznym i emotywnym, wysoce istotna pozostaje zauważalna tendencja autora do rejestrowania nie tylko typowych frazeologizmów, lecz także innych utartych połączeń wyrazowych. Skorupka zaprezentował szeroką koncepcję frazeologii, uwzględniającą nie tylko klasyczne związki frazeologiczne, lecz także konstrukcje porównawcze, sentencje, przysłowia, stereotypy oraz utarte zwroty języka mówionego.

Jednostki wielowyrazowe, jak można się spodziewać, nie funkcjonują w słowniku jako samodzielne hasła główne, lecz stanowią rozwinięcie głównych artykułów hasłowych, często w sposób rozbudowany i szczegółowy. Przykładowo, hasło *chyba* zawiera, m.in. formę *chyba tak* (wyrażającą pewność), a także charakterystyczny dla języka mówionego fatyzm *chyba, że...* (sygnalizujący pewne ograniczenie). Wśród pozostałych haseł odnaleźć można szereg wyrażen o charakterze emotywnym, takich jak *co za* (*co to za?*); *co tam*; *co mi tam*; *jeszcze czego!* czy *co z tego*. Na uwagę zasługuje również stosowanie kwalifikatorów potoczności czy ironiczności, co należy uznać za zabieg wyjątkowy, ponieważ w tamtym okresie tradycyjna

leksykografia unikała rejestracji jednostek takiego rejestru. Z tej perspektywy przedstawioną pracę można traktować jako próbę wskazania potrzeby systematycznej dokumentacji tego typu jednostek, potwierdzenia ich rozpowszechnienia w języku oraz znaczenia komunikacyjnego, skoro zostały one uwzględnione również w klasycznym opisie frazeologicznym.

Należy podkreślić, że koncepcja Doroszewskiego została w zmodyfikowanej formie rozwinięta w trzypięciotomowym *Słowniku języka polskiego PWN* pod redakcją Mieczysława Szymczaka (1978–1981). Dzieło to, choć powstawało w innej rzeczywistości społeczno-politycznej, nawiązywało pośrednio do modelu Doroszewskiego poprzez, m.in. silnie zaznaczoną funkcję kulturową i normatywną. Należy jednak podkreślić, że zdecydowanie wykazywało większą otwartość na nowsze tendencje w języku ogólnym, co potwierdza obecność form potocznych czy kolokwialnych. Pomimo szerokiej krytyki za niespójność metodologiczną, nadmiar terminologii specjalistycznej czy ograniczone uwzględnianie języka potocznego, słownik zawiera liczne frazeologizmy i jednostki wielowyrazowe, takie jak *jak nic; kto jak kto; goły jak święty turecki* czy *jak gdyby nigdy nic*, choć nie funkcjonują one jako odrębne artykuły hasłowe. W obrębie haseł głównych pojawiają się również wyrażenia o funkcji fatycznej i ematywnej – np. *jak to; a jakże; jazda stąd!; a nie mówiłem?; po jaką cholere?* – wplecione w materiał ilustracyjny lub kolokacyjny. W słowniku w sposób rozproszony pojawiają się elementy języka mówionego, a kwalifikatory stylistyczne stosowane są dosyć rzadko. Mimo braku wyraźnych założeń teoretycznych dotyczących tego rodzaju jednostek, autorzy tego słownika w sposób ostrożny i ograniczony zaczynają sygnalizować obecność wielowyrazowych reproduktów w ówczesnej polszczyźnie.

Kolejnym znaczącą postacią w leksykografii polskiej jest Maciej Grochowski, który zaproponował alternatywną wobec tradycji Doroszewskiego koncepcję opracowania słownika. W 1982 roku w swojej pracy *Zarys leksykologii i leksykografii* oraz późniejszych publikacjach zaproponował on model słownika opartego na współczesnych osiągnięciach semantyki, gramatyki i różnych dziedzin lingwistyki. Grochowski krytykował dotychczasowy, niespójny metodologicznie charakter słowników, odchodząc tym samym od definicji intuicyjnych i encyklopedycznych. Postulował również o wprowadzenie rygorystycznych kryteriów doboru haseł i podhaseł oraz unikanie znaczeń specjalistycznych wymagających specjalistycznego opisu metajęzykowego. Praktyczną realizacją jego koncepcji stał się *Słownik polskich przekleństw i wulgaryzmów* z 1995 roku, w którym choć nie wyodrębniono odrębnych haseł wielowyrazowych, wiele jednostek ematywnych pojawia się w analizach jako przykłady użycia, np. *do licha; Boże święty; w mordę jeża; o rety; cholera jasna*. Słownik ten był przełomowy głównie ze względu na uwzględnienie słownictwa wulgarne i potoczne,

stanowiąc realizację postulatu opisu realnego języka z uwzględnieniem w ograniczonym zakresie także jednostek wielowyrazowych. W tym samym okresie, a więc w latach 80., istotny głos w dyskusjach metodologicznych zabierali również Zygmunt Saloni i jego współpracownicy, których propozycje stanowiły wyraźną przeciwwagę wobec stanowiska Grochowskiego. Saloni postulował uzupełnienie opisu leksykalnego o dane gramatyczne, większą precyzję w klasyfikacji składniowej oraz eliminację terminów specjalistycznych i rzadko używanych. Główne postulaty jego zespołu dotyczyły integracji informacji o budowie składniowej i fleksyjnej jednostek oraz dostosowania słownika ogólnego do rzeczywistych potrzeb użytkownika. Równoległe do tych ujęć rozwijała się również koncepcja Andrzeja Bogusławskiego, szczególnie istotna z punktu widzenia badań nad odtwarzalnymi jednostkami języka mówionego. Autor ten jako jeden z pierwszych konsekwentnie opowiadał się za systemowym wyodrębnianiem odtwarzalnych wielowyrazowych jednostek jako pełnoprawnych obiektów leksykograficznych, wymagających odrębnego opisu.

Wśród polskich leksykografów końca XX wieku na uwagę zasługuje również Halina Zgółkowa, kierująca zespołem *Praktycznego słownika współczesnej polszczyzny* (1994–2005). Jej projekt, skonstruowany jako słownik przyjazny użytkownikowi, miał charakter ściśle praktyczny i odchodził od sformalizowanej teorii na rzecz konkretnych rozwiązań redakcyjnych. Autorka dążyła do uproszczenia struktury haseł, rezygnując, m.in. ze skrótów, kwalifikatorów i materiału dokumentacyjnego, a także grupując wszystkie informacje o wyrazie w jednym miejscu. Warto nadmienić, iż zawarte w pracy artykuły hasłowe cechuje obszerna, przejrzysta forma, a szczególną uwagę poświęcono szerokiemu repertuariowi kolokacji o najwyższej frekwencji użycia.

Zgółkowa w swojej refleksji pośrednio zbliżyła się do założeń Tadeusza Piotrowskiego, postulującego w latach 90. XX wieku odejście od słownika jako modelu kompetencji językowej leksykografa na rzecz słownika jako narzędzia poznania użytkownika. Autor podkreślał, że „użytkownicy traktują słownik jako tekst, który informuje ich o tym, czego nie wiedzą, nie zaś o tym, co już wiedzą” (Piotrowski 1994: 72). W jego ujęciu konieczne staje się myślenie o potrzebach odbiorców i tworzenie słowników-prototypów, dostosowanych do oczekiwań konkretnej wspólnoty językowej.

W leksykografii tego okresu istotne miejsce zajmuje również *Słownik współczesnego języka polskiego* pod redakcją Bogusława Dunaja (1996), który powstawał niezależnie od wcześniejszych słowników normatywnych i jednocześnie stanowił próbę rejestracji żywego zasobu języka – także w jego odmianie potocznej. Szczególnie innowacyjny pozostaje sposób ujęcia jednostek wielowyrazowych, pojawiających się zarówno w obrębie haseł

jednowyrazowych, jak i jako samodzielne artykuły hasłowe, opatrzone definicjami, kwalifikatorami i przykładami użycia. Spośród licznych przykładów warto wymienić takie konstrukcje jak: *proszę ja ciebie; ależ oczywiście; daj spokój; a nie mówiłem; halo; no to ładnie; a masz; no i co z tego* – pełniące przede wszystkim funkcje fatyczne i ematywne. Oprócz nich słownik zawiera też tradycyjne frazeologizmy, np. *grać pierwsze skrzypce; mieć serce na dłoni; żyć jak pies z kotem*. Takie ujęcie dowodzi, że autorzy słownika dostrzegali potrzebę opisu języka codziennego, bardziej zbliżonego do realnej praktyki komunikacyjnej niż do wzorca normatywnego.

W nawiązaniu do powyższych koncepcji warto również przywołać *Słownik polszczyzny potocznej* autorstwa Janusza Anusiewicza i Jacka Skawińskiego z 1996 roku, który jako jedno z nielicznych opracowań poświęcone jest leksyce potocznej. Słownik liczy około 5000 haseł i został opracowany na podstawie różnorodnych źródeł: tekstów mówionych, literatury, prasy, filmów oraz zasłyszanych wypowiedzi. Artykuły hasłowe zostały wzbogacone o krótkie definicje oraz realne przykłady użycia, a styl objaśnień dostosowano do potocznego charakteru materiału. Na szczególną uwagę z perspektywy prowadzonych badań zasługuje wysoka frekwencja jednostek wielowyrazowych, wśród których można zaobserwować, m.in.: emotywy (np. *i po strachu!; nie ma siły!; ani rusz!*). Obecność takich konstrukcji dowodzi, że słownik ten nie tylko wykracza poza ówczesny tradycyjny opis słownikowy, ale posiada sporą wartość pragmatyczną dla użytkownika.

Za umowny początek nowego etapu polskiej leksykografii jednojęzycznej XXI wieku można uznać wydanie w 2000 roku *Innego słownika języka polskiego PWN* pod redakcją Mirosława Bańki, który zrywa z dotychczasową tradycją opisu leksykograficznego. Konstrukcja tego słownika jest oparta na danych empirycznych, analizie użycia i klarownym opisie definicyjnym, wprowadzając nowe rozwiązania zarówno w warstwie metodologicznej, jak i stylistycznej. Proste i naturalne definicje są często osadzone w kontekstach narracyjnych lub sytuacyjnych (np. *Nie czekaj z obiadem!; Proszę pozdrowić ode mnie pana Mariusza*), co nadaje im autentyczności i sprawia, że przypominają one wypowiedzi dialogowe. Właśnie w takich kontekstach szczególnie wyraźnie można zauważyć obecność licznych jednostek wielowyrazowych, jak *muszę lecieć; pa pa; jak sobie chcesz*, które wcześniej rzadko rejestrowały słowniki o charakterze ogólnym. Zestawienie to uzupełniają liczne przysłowia, idiomy i wyrażenia metaforyczne (*ryba psuje się od głowy; chcieć to móc; mieć świeka w głowie; wylać kubel zimnej wody*), które nie są traktowane marginalnie, lecz stanowią integralną część hasłowych artykułów. Konstrukcje te nie tylko wzbogacają definicje o wymiar naturalności i realności użycia, lecz także dają użytkownikowi bezpośredni wgląd w

autentyczne sytuacje komunikacyjne. Nowatorstwo *ISJP PWN* jest również widoczne w dwukolumnowej strukturze artykułu hasłowego, z których pierwsza podaje informacje podstawowe, natomiast druga skupia się na szczegółowych danych, co pozwala dostosować słownik do potrzeb zarówno przeciętnego, jak i zaawansowanego użytkownika. Warto jednak zaznaczyć, że pomimo wyraźnego zwrotu w kierunku języka mówionego, *ISJP* pozostaje opracowaniem selektywnym i nie w pełni kompletnym z perspektywy badań nad wielowyrazowymi emotywami i fatyzmami. Choć formy tego typu pojawiają się w słowniku stosunkowo często, nie są traktowane jako odrębna kategoria językowa – brak im zarówno systematycznego opisu, jak i graficznego wyróżnienia w strukturze artykułu hasłowego. W rezultacie wyrażenia takie jak *jak Bóg miły; Boże; na Boga* czy *pa, pa* występują w słowniku w sposób rozproszony, co uniemożliwia ich postrzeganie jako konsekwentnie opracowanego zbioru materiału do analiz jednostek wielowyrazowych. Z tej perspektywy *ISJP PWN*, mimo swojego niewątpliwie innowacyjnego charakteru i przełomu w leksykografii jednojęzycznej, stanowi raczej sygnał konieczności podjęcia dalszych badań w tym obszarze niż ich pełną realizację.

Kolejną istotną publikacją, wydaną w 2003 roku przez Wydawnictwo Naukowe PWN, był czterotomowy *Uniwersalny słownik języka polskiego* (dalej: *USJP*) pod redakcją Stanisława Dubisza, zawierający około 100 000 haseł i stanowiący kontynuację oraz rozbudowę wspomnianego *Słownika języka polskiego* pod redakcją Mieczysława Szymczaka. Zgodnie z tytułem, celem publikacji było opracowanie tzw. *słownika uniwersalnego*, łączącego w sobie różne typy danych leksykograficznych, np. etymologiczne, gramatyczne, składniowe, frazeologiczne, pragmatyczne czy stylistyczne. Choć struktura *USJP* w dużej mierze pozostała wierna klasycznym wzorcom, wprowadzone zostały istotne zmiany względem słownika Szymczaka, tj. poszerzono zasób haseł i rozbudowano repertuar kwalifikatorów stylistycznych, a informacje gramatyczne umieszczono na końcu artykułu hasłowego. Warto zauważyć, że w słowniku tym skoncentrowano się głównie na słownictwie drugiej połowy XX wieku, co przejawia się w rejestracji leksyki ogólnej i potocznej przy jednoczesnym pomijaniu wyrazów archaicznych i specjalistycznych. Z perspektywy badań nad reproduktami kluczowe znaczenie ma częściowe udostępnienie *USJP* w wersji elektronicznej na stronie sjp.pwn.pl, która łączy materiały z kilku słowników PWN (w tym *USJP*, *WSJP* i *ISJP*), prezentując informacje o hasłach w sposób dość niespójny i niejednorodny. Jeśli chodzi o rejestrację jednostek wielowyrazowych, odnotowano znaczną liczbę idiomów, przysłów i frazeologizmów, posiadających samodzielne artykuły hasłowe, opatrzone definicjami, kwalifikatorami i przykładami użycia. Natomiast w odniesieniu do analizowanych fatyzmów i emotywów

sytuacja przedstawia się mniej przejrzysto. Pomimo faktu, iż niektóre z nich (np. *a nie mówiłem?*; *coś ty!*; *proszę cię!*) można wyszukać w systemie sjp.pwn.pl, to nie zawsze funkcjonują jako odrębne hasła, lecz często pojawiają się wyłącznie jako przykłady użycia w ramach haseł głównych (np. *coś*, *prosić*). Ich opis jest zazwyczaj dość ubogi, pozbawiony definicji oraz kwalifikatorów stylistycznych, a same jednostki nie są graficznie wyróżnione ani traktowane jako odrębne obiekty opisu leksykograficznego. Podsumowując, należy zaznaczyć, że choć *USJP* powstał w oparciu o nowoczesne założenia i funkcjonuje również w wersji elektronicznej, nie prezentuje on pogłębionego opisu wielowyrazowych jednostek języka mówionego, zwłaszcza fatycznych czy emotywnych. Fakt ten nabiera szczególnego znaczenia w kontekście rosnącej roli słowników online w codziennej praktyce użytkowników, zwłaszcza że sjp.pwn.pl należy współcześnie do najczęściej wykorzystywanych opracowań leksykograficznych w polskim internecie.

Ostatnim z analizowanych w tym miejscu współczesnych słowników, który wydaje się być najbardziej obiecującym projektem polskiej leksykografii jednojęzycznej, jest rozwijany od 2006 roku *Wielki słownik języka polskiego PAN*, dostępny w wersji elektronicznej pod adresem wsjp.pl. Opracowanie to, tworzone pod kierunkiem Piotra Źmigrodzkiego, zostało zaprojektowane jako nowoczesne narzędzie elektroniczne, którego podstawowym celem jest rejestracja rzeczywistego i aktualnego użycia języka polskiego w różnych jego rejestrach przy wykorzystaniu danych korpusowych. Autorzy słownika kierują się zasadami deskryptywności i dokumentacyjności, dążąc do możliwie najwierniejszego i najbardziej precyzyjnego opisu faktycznego stanu języka, co sprawia, że *WSJP* stanowi modelowy przykład nowoczesnej leksykografii, ukierunkowanej na empiryczne zarejestrowanie współczesnej polszczyzny. Należy zwrócić szczególną uwagę na innowacyjne ujęcie jednostek wielowyrazowych, które są prezentowane jako samodzielne, pełnoprawne hasła oraz są wyposażone w wiele informacji leksykograficznych. W słowniku znajdziemy wiele wyrażen fatycznych, takich jak *a więc*; *tak czy owak*; *jakby tego było mało* oraz emotywnych, na przykład *a nie mówiłem*; *ależ oczywiście*; *daj spokój*; *no to pięknie*; *no i co z tego*. Jednostki te są opatrzone jasnymi definicjami, kwalifikatorami stylistycznymi i tematycznymi, a także wariantami i przykładami użycia, zaczerpniętymi z języka mówionego. Słownik ujmuje tego typu wyrażenia jako odtwarzalne całości o określonej funkcji pragmatycznej, co niewątpliwie wpisuje się w nowoczesne tendencje leksykografii korpusowej. Warto wspomnieć również o przejrzystym i funkcjonalnym interfejsie oraz modułowej strukturze haseł, która umożliwia szybki dostęp do potrzebnych informacji, a szczegółowy opis założeń i budowy słownika został przedstawiony w monografii *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*.

W kontekście badań nad fatycznymi i emotywnymi reproduktami WSJP okazuje się narzędziem szczególnie przydatnym, ponieważ stanowi nie tylko źródło informacji o jednostkach jednowyrazowych, które dominują w jego strukturze, lecz także uwzględnia liczne wielowyrazowce jako samodzielne hasła. Traktowanie danych jednostek jako pełnoprawnych składników współczesnej polszczyzny, połączone z precyzyjnym opisem i uwzględnieniem języka mówionego, świadczy o wyraźnym zwrocie ku nowemu modelowi leksykografii, który zdaje się poszerzać granice tradycyjnego opisu o jednostki wielowyrazowe.

Jak wykazała powyższa analiza, podejście do jednostek wielowyrazowych w polskiej leksykografii jednojęzycznej przeszło zauważalną ewolucję – od ich marginalizacji w modelu normatywnym po coraz bardziej systematyczne ujmowanie w najnowszych słownikach opartych na danych korpusowych. Choć w dalszym ciągu rejestrowane są hasła jednowyrazowe, współczesne tendencje wyraźnie zmierzają w stronę opisu także reproduktów wielowyrazowych, traktowanych jako pełnoprawne jednostki systemu leksykalnego. Wynika to zapewne z rosnącej świadomości potrzeby kodyfikacji i konsekwentnego opisu wyrażen wielowyrazowych, których rola w codziennej komunikacji zaczyna być coraz wyraźniej dostrzegana. Zagadnienie to niewątpliwie wymaga odrębnego, pogłębionego opracowania, jednak już na obecnym etapie można dostrzec wyraźne przesunięcie ku nowej teorii polskiej leksykografii. W dalszej części rozdziału podjęta zostanie próba analizy stanu i tendencji w rosyjskich słownikach jednojęzycznych, ze szczególnym uwzględnieniem sposobu rejestracji jednostek wielowyrazowych o funkcji fatycznej i ematywnej.

1.7 Rejestracja jednostek fatycznych i emotywnych we współczesnych rosyjskich słownikach jednojęzycznych

Rewolucja październikowa oraz reforma ortograficzna z 1917 roku zapoczątkowały głębokie przemiany w systemie języka rosyjskiego, które wymusiły proces dezaktualizacji dotychczasowych słowników i opracowania nowych. Rosyjska leksykografia jednojęzyczna XX i XXI wieku rozwijała się w oparciu o kilka równoległych nurtów: normatywny, opisowy oraz funkcjonalno-komunikacyjny. Każdy z nich odzwierciedlał inne podejście do języka, jego roli społecznej i metod opisu.

Wśród kluczowych postaci tego okresu szczególne miejsce zajmuje Lew Szczerba (1880–1944), uznawany za twórcę nowoczesnej teorii leksykografii jako samodzielnej dziedziny lingwistyki. W swojej fundamentalnej rozprawie *Опыт общей теории лексикографии* (1940) sformułował podstawy klasyfikacji słowników oraz zaproponował

ujęcie słownika jako naukowego modelu systemu językowego. Postulował również tworzenie słowników, uwzględniających nie tylko struktury semantyczne i gramatyczne, ale także stylistyczne zróżnicowanie leksyki oraz kontekstowe użycie języka. Jako jeden z pierwszych postulował koncepcję systemu słowników narodowego języka rosyjskiego i ideograficznego porządkowania leksyki na podstawie relacji znaczeniowych. W tym nurcie powstał między innymi czterotomowy *Толковый словарь русского языка* pod redakcją Dmitrija Uszakowa, nad którym prace trwały w latach 1928–1940. Opracowanie to, liczące ponad 85 000 jednostek, jest normatywne, a szczególną uwagę poświęcono zasadom doboru leksyki, informacji fleksyjnej, fonetycznej oraz pragmatycznej. Warto podkreślić rozbudowany system kwalifikatorów stylistycznych, uwzględniających również odniesienia do języka mówionego (np. kwalifikatory *разг.*, *простореч.*). Obecność wyrażen potocznych, ironicznych i ekspresywnych świadczy o dążeniu autorów do wiernego odzwierciedlenia rzeczywistego stanu języka – nie tylko literackiego, ale również potocznego. Wspomniany autor do dziś uznawany jest za jednego z najwybitniejszych leksykografów rosyjskich, a opracowany przez niego słownik uchodzi za wzorcowy przykład prawidłowego opisu leksykograficznego w Rosji. Choć nie zawiera on samodzielnych jednostek wielowyrazowych o funkcji fatycznej czy emotywnej, notacja elementów języka mówionego stanowi istotne świadectwo rejestracji żywej, dynamicznej odmiany języka w ujęciu słownikowym. Utrata kartoteki tego słownika po II wojnie światowej sprawiła, że rosyjska leksykografia stanęła przed ponownym wyzwaniem zebrania reprezentatywnego materiału leksykalnego, co doprowadziło najpierw do powstania pracy *Малый академический словарь*, a następnie do opracowania projektu *Большой академический словарь русского языка* (БАС), którego celem było odzwierciedlenie pełnego obrazu współczesnego języka rosyjskiego. Opracowanie to było redagowane w latach 1999–2017 przez Instytut Języka Rosyjskiego im. W. W. Winogradowa Rosyjskiej Akademii Nauk pod kierownictwem naukowym Jurija Apresjana. Słownik, który ukazał się w trzydziestu tomach i liczy około 150 000 haseł, rejestruje nie tylko leksykę literacką, lecz także słownictwo potoczne, specjalistyczne, regionalne i ekspresywne. Uwzględnienie poszczególnych jednostek uzależniono od ich poświadczenia w źródłach pisanych. Struktura artykułu hasłowego zawiera szczegółowy opis gramatyczny i semantyczny, kwalifikatory stylistyczne, segmenty frazeologiczne oraz liczne cytaty literackie. Zawarte w słowniku jednostki wielowyrazowe (frazologizmy, kolokacje, konstrukcje składniowe) pojawiają się zarówno jako odrębne hasła, jak i jako uzupełnienia haseł głównych, a ich udział szacuje się na około 5–7% całkowitej zawartości słownika. W pracy znajdziemy również pojedyncze jednostki o charakterze fatycznym i emotywnym, jak *Ну и ну!*; *Вот так да!*; *Да ну!*; *Ай да!*, które opatrzone są

kwalifikatorami typu *разг., эмоц., экспр.* Istotne miejsce zajmują także reproduktory fatyczne, takie jak *по правде говоря; между прочим; как говорится; надо сказать*, określane zwykle jako *вводные выражения* i ilustrowane przykładami kontekstowymi. Pomimo faktu, iż udział jednostek fatycznych i emotywnych pozostaje stosunkowo niewielki (szacunkowo 1–1,5%), ich obecność wskazuje na świadomy kierunek obrany przez redaktorów, polegający na rejestracji różnorodnych funkcji komunikacyjnych języka, od stylu wysokiego po język potoczny. Można zatem stwierdzić, iż dane opracowanie stanowi wzorcowy przykład nowoczesnej leksykografii akademickiej, opartej o korpus i wyselekcjonowaną dokumentację, aczkolwiek kodyfikacja jednostek fatycznych i emotywnych ma w nim charakter raczej fragmentaryczny. Ujęcie tego typu jednostek sygnalizuje dopiero początek nowego etapu w opisie wyrażen pragmatycznych w leksykografii języka rosyjskiego, który dopiero w kolejnych dziesięcioleciach będzie podlegał znaczącym zmianom.

W XX wieku w rosyjskiej leksykografii uformował się trójstopniowy model słowników jednojęzycznych, w ramach którego zaczęto wyróżniać:

- 1) *Słownik jednotomowy* – przeznaczony dla przeciętnego użytkownika, prezentujący podstawowy zasób leksyki i frazeologii języka rosyjskiego;
- 2) *Słownik czterotomowy*, ukazujący stylistyczne zróżnicowanie języka literackiego;
- 3) *Słownik wielotomowy* o charakterze specjalistycznym, uwzględniający możliwie najszersze spektrum literackiej odmiany języka i adresowany głównie do specjalistów.

Należy w tym miejscu wspomnieć o jednym z najpopularniejszych słowników jednotomowych XX wieku, a mianowicie *Толковый словарь русского языка* autorstwa Sergieja Ożegowa, którego pierwsze wydanie ukazało się w 1949 roku. Głównym celem danego opracowania było upowszechnienie kultury językowej w możliwie najszerszych kręgach odbiorców, co znalazło odzwierciedlenie w jego zwartej strukturze oraz charakterze funkcjonalnym. W odróżnieniu od swoich poprzedników, zastosowano ograniczony opis semantyczny i stylistyczny, a przykłady użycia w zdaniach miały w sposób lakoniczny ilustrować poprawne użycia konkretnej jednostki. Całkowicie wyeliminowano wszelką terminologię naukową, regionalizmy czy archaizmy, co sprawiło, że słownik zyskał status praktycznego narzędzia przeznaczonego dla szerokiego grona odbiorców. Po śmierci autora w 1964 roku redakcję objęła Natalia Szwedowa, która stopniowo modyfikowała słownik i uzupełniała jego treść o nowsze jednostki zgodne z ewoluującymi normami języka rosyjskiego. W najnowszych wydaniach słownik liczy już około 72 500 jednostek jednowyrazowych oraz blisko 7 500 wyrażen frazeologicznych, a jego zawartość jest systematycznie aktualizowana i dostosowywana do potrzeb współczesnych użytkowników. Z perspektywy badań nad

reproduktami fatycznymi i emotywnymi warto podkreślić, że mimo wyraźnie normatywnego charakteru słownika, uwzględniono w nim znaczną liczbę wyrażen charakterystycznych dla języka mówionego. Wśród nich można odnaleźć emotywy, takie jak *Да ну!; Ты что!; Ну и ну!; Ты что, с ума сошел?*, oraz fatyzmy typu *Между прочим; По правде говоря; Как говорится*. Jednostki te pojawiają się głównie jako uzupełnienia haseł głównych, a także w przykładach użycia i pełnią funkcje łącznikowe, sygnalizując zmianę tematu lub komentarz osoby mówiącej. Pomimo stosunkowo niewielkiego udziału jednostek wielowyrazowych w strukturze słownika (około 3% całego zasobu), ich obecność sygnalizuje wyraźny zwrot w ówczesnej leksykografii rosyjskiej w kierunku rejestracji utrwalonych jednostek języka mówionego oraz ich funkcji pragmatycznych.

Jednym z najważniejszych współczesnych słowników jednojęzycznych języka rosyjskiego pozostaje *Большой толковый словарь русского языка* pod redakcją Sergieja Aleksandrowicza Kuzniecowa z 1998 roku, łączący tradycyjne podejście opisowe z elementami normatywnymi. Opracowanie to miało na celu odzwierciedlenie rzeczywistego stanu języka rosyjskiego poprzez ukazanie słownictwa literackiego, potocznego i technicznego w różnych kontekstach komunikacyjnych. W rezultacie powstało dzieło o charakterze uniwersalnym, rejestrujące około 120 000 jednostek leksykalnych, przeznaczone zarówno dla badaczy, nauczycieli czy tłumaczy, jak i dla przeciętnego użytkownika języka. Struktura artykułu hasłowego w danym opracowaniu jest przejrzysta i szczegółowa i zawiera, m.in.: formę podstawową z akcentem, klasyfikację morfologiczną, kwalifikatory stylistyczne (*разг., экспр., книжн.*), definicje znaczeń oraz przykłady użycia, w tym frazeologizmy. Jednostki wielowyrazowe stanowią integralną część artykułów hasłowych i uwzględniają zarówno klasyczne przykłady idiomów (*лечь в воду; дело в шляпе*), kolokacji (*принять участие; оказать помощь*), jak i przysłów (*время — деньги; нет дыма без огня*), a ich udział szacuje się na około 5–7% całkowitej zawartości słownika. Istotny pozostaje sposób ujęcia jednostek wielowyrazowych o funkcji fatycznej i ematywnej, które nie są, co prawda, wyodrębnione jako samodzielne hasła, lecz pojawiają się stosunkowo często w przykładach użycia oraz w sekcjach frazeologicznych.

Wśród wyrażen fatycznych znajdują się, m.in.: *на самом деле; по моему мнению; я думаю; мне кажется*, pełniące funkcje metatekstowe, organizujące wypowiedź lub wyrażające subiektywną opinię mówiącego.

Reprodukty ematywne, takie jak *вот так да!; ого!; а что?; что за...*, występują głównie przy hasłach związanych z ekspresją i są oznaczone kwalifikatorami *разг.* lub *экспр.*, a także ilustrowane przykładami dialogów, co świadczy o próbie uchwycenia natury języka

mówionego. W tym kontekście dane opracowanie należy uznać za jedno z najpełniejszych współczesnych źródeł wiedzy o rzeczywistym funkcjonowaniu języka rosyjskiego, uwzględniające zarówno jego odmianę literacką, jak i potoczną. Jego szczególna wartość dla badań nad jednostkami wielowyrazowymi, w tym wyrażeniami fatycznymi i emotywnymi, wynika z bogatego materiału ilustracyjnego, mimo że jednostki te nie są traktowane jako odrębne obiekty opisu leksykograficznego.

Należy w tym miejscu wspomnieć również o innowacyjnym słowniku przełomu XX i XXI wieku, tj. pracy *Новый толково-словообразовательный словарь русского языка* autorstwa Tatiany Jefremowej opublikowanej w 2000 roku. Opracowanie, liczące ponad 120 000 haseł, wyróżnia się otwartością na rejestrację jednostek funkcjonalnych, dialogowych i metatekstowych, co czyni go szczególnie cennym źródłem dla badań nad wielowyrazowymi jednostkami języka mówionego. Autorka stworzyła dzieło zorientowane nie tylko na użytkownika akademickiego, lecz także na szeroki krąg odbiorców, takich jak nauczyciele, tłumacze oraz studenci. Jego wersja elektroniczna jest obecnie dostępna na wielu platformach internetowych. Na tle innych opracowań szczególnie odznacza się sposób traktowania jednostek wielowyrazowych, które nie tylko są szeroko rejestrowane, lecz także posiadają samodzielne artykuły hasłowe. Znalazły się wśród nich związki składniowe, wyrażenia modalne, konstrukcje dialogowe i utarte frazy, takie jak *а если бы*; *а ну как*; *а то как же*; *по сути дела*; *по большому счёту*; *как бы то ни было*. Szczególne miejsce zajmują również jednostki fatyczne i ematywne, które otrzymują status pełnoprawnych artykułów hasłowych i są opatrzone informacją o ich funkcji komunikacyjnej (np. wyrażenie protestu, aprobaty, zaskoczenia, ironii czy zachwyty). Wśród wyrażen fatycznych znajdziemy, m.in. *так сказать*; *надо сказать*; *а если бы*, a do jednostek emotywnych można odnieść *Что вы!*; *Ну и ну!*; *Вот оно что!*; *И что же?*. Szacunkowy udział jednostek wielowyrazowych jako samodzielnych haseł w tym słowniku wynosi około 17%, z czego 2–3% to wyrażenia fatyczne i ematywne. Dzięki ich systematycznej rejestracji słownik ten nie tylko dostarcza rozbudowanych informacji o znaczeniach i użyciu słów, lecz także stanowi próbę uchwycenia żywego języka mówionego w jego wymiarze funkcjonalnym, emocjonalnym i komunikacyjnym.

Pracą uznawaną za jedno z lepszych opracowań rosyjskiej leksykografii frazeologicznej pozostaje *Фразеологический словарь русского литературного языка* autorstwa Aleksandra Fiodorowa, którego trzecie, rozszerzone wydanie z 2008 roku liczy około 13 000 jednostek frazeologicznych. Słownik ten stanowi próbę kompleksowego ujęcia utrwalonych połączeń leksykalnych, jednak mimo zadeklarowanego celu rejestracji pełnego spektrum frazeologii literackiej, w rzeczywistości koncentruje się głównie na tradycyjnych idiomach, kolokacjach i

przysłowiach, traktując jednostki fatyczne i ematywne w sposób wybiórczy. Na tle innych opracowań frazeologicznych słownik ten wyróżnia się jednak pewnymi elementami, takimi jak uzupełnianie definicji kwalifikatorami stylistycznymi (np. *разг.*, *экспр.*, *книжн.*) oraz wyodrębnianiem wielowyrazowych reproduktów fatycznych i ematywnych jako odrębnych artykułów hasłowych, np. *ну и ну!*; *да ну!*; *ах ты!*; *между прочим*; *так сказать*; *а вот*; *а вот и нет*; *а вот на тебе!*; *а поди ж ты*; *а то как же*; *а не то*; *а между тем*; *а то бы вот*. Jednostki te przedstawione są w sposób tradycyjny, z naciskiem na ich cechy ekspresywne oraz wyrażanie zaskoczenia, dezaprobaty lub potwierdzenia, przy czym użytkownik otrzymuje jedynie zwięzły opis ich znaczenia, bez pogłębionej analizy funkcji pragmatycznych. Często opatrzone są one rozbudowanymi przykładami użycia w dialogu, jednak mimo takiej dokumentacji autor nie wyodrębnia ich jako odrębnej kategorii, co może utrudniać ich systematyczną analizę. Słownik Fiodorowa należy zatem uznać za źródło wartościowe, choć w dużym stopniu ograniczone ze względu na skoncentrowanie się na dokumentacji klasycznego zasobu frazeologicznego, a nie na szczegółowym ujęciu języka mówionego, w tym fatyzmów i ematywów. Mimo rejestracji dobrze opracowanych przykładów jednostek wielowyrazowych, są to głównie najbardziej znane i powielane formuły, co sprawia, że słownik ten może stanowić jedynie uzupełnienie badań nad jednostkami fatycznymi i ematywnymi, a nie podstawę ich analizy.

Do wyjątkowych opracowań współczesnej rosyjskiej leksykografii odnieść należy również *Толковый словарь русской разговорной речи* pod redakcją Leonida Krysina, który jest publikowany od 2014 roku. Słownik ten stanowi udaną próbę systematycznej dokumentacji języka potocznego, który traktowany jest na równi z językiem literackim i uznawany za pełnoprawną odmianę języka ogólnego, zasługującą na szczegółowy opis. Opracowanie to ma charakter deskryptywny i różnicujący, rejestrując jedynie jednostki typowe dla mowy potocznej, używanej głównie przez wykształconych mieszkańców rosyjskich miast, a nieobecne w normie pisemnej języka literackiego. Jedną z cech wyróżniających słownik jest wysoka obecność jednostek wielowyrazowych, które choć nie występują jako samodzielne hasła, to są wyodrębniane w ramach haseł jednowyrazowych. Pojawiają się one między innymi w sekcjach frazeologicznych (PHRAS), syntaktycznych (SYNT) oraz pragmatycznych (PRAGM) i są graficznie wyróżnione pogrubioną czcionką. Wśród tych wyrażen znajdują się zarówno idiomy, mniej używane kolokacje, składniki dyskursywne, struktury dialogowe, jak również liczne fatyzmy i ematywy. Ilustracyjnie można przytoczyć liczne reproduktów towarzyszące artykułowi „так”, np. *и так и этак*; *как бы не так!*; *как (же это) так!*; *не так чтобы (очень уж)*; *вот так*; *так, да не так*. Wyrażenia te nierzadko pełnią funkcje

emotywne (wyrażenie zaskoczenia, rezygnacji, zadowolenia) lub fatyczne (regulowanie toku wypowiedzi, skupienie uwagi odbiorcy), a ich obecność w haśle jednowyrazowym nie tylko świadczy o rzeczywistym funkcjonowaniu tych struktur w języku, ale także o świadomości autorów dotyczącej ich znaczenia komunikacyjnego. Przykłady takie jak *здрасьте, я ваша тѣтя* (wyrażenie zaskoczenia i dezaprobaty), *мѹна того; всѣ мѹн-мон; будь спок; ну и ну!* dowodzi rejestracji ekspresywnych formuł komunikacyjnych, które cechują się wysoką frekwencją użycia w języku mówionym. Można zatem podsumować, iż słownik Krysina stanowi niewątpliwie cenne źródło dla badań nad językiem mówionym ze szczególnym uwzględnieniem wielowyrazowych jednostek fatycznych i emotywnych. Jego przejrzysta struktura funkcjonalna oraz systematyczne uwzględnianie jednostek wielowyrazowych w wysokim stopniu odpowiadają aktualnym potrzebom leksykografii pragmatycznej i wpisują się w kierunek poszukiwań nastawionych na analizę wielowyrazowców i ich funkcji komunikacyjnych.

Na krótkie omówienie zasługuje również otwarta i dynamicznie rozwijana kategoria *Russian multiword terms*, opracowywana w ramach słownika *Wiktionary*, która mimo braku statusu słownika akademickiego stanowi jedno z najbogatszych elektronicznych źródeł rejestrujących współczesne jednostki wielowyrazowe języka rosyjskiego. Zgromadzone tam hasła uwzględniają w głównej mierze frazeologizmy, przysłowia, idiomy, ale także wyrażenia fatyczne i ekspresywne, których próżno szukać w porównywalnej liczbie w innych opracowaniach rosyjskiej leksykografii jednojęzycznej. Ich opisy są zazwyczaj bardzo uproszczone, często ograniczają się do krótkiej definicji i przekładu na język angielski, jednak pomimo tej wady, ów zbiór jest wysoce przydatny do identyfikowania reproduktów języka mówionego oraz weryfikowania ich znaczenia w innych źródłach. Szczególnie liczną grupę stanowią wyrażenia o silnym nacechowaniu emocjonalnym, takie jak *в чем дело?*; *вру больше!*; *ещѣ чего!*; *аѹ да мы!* czy *вот это да*, które należy traktować jako wyrażenia reakcyjne. Choć opisy jednostek bywają nierówne pod względem szczegółowości i opracowania, to jednak ich liczba oraz aktualność zgromadzonych form sprawiają, że kategoria ta stanowi wartościowe uzupełnienie materiału słownikowego, szczególnie przydatne w badaniach nad fatyzmami i emotywami. Na tle innych języków słowiańskich warto podkreślić, że tak rozbudowanego i zróżnicowanego zasobu, zawierającego wielowyrazowe konstrukcje języka mówionego, nie znajdziemy ani w polskiej ani czeskiej wersji *Wiktionary*. Brakuje tam porównywalnie systematycznej kategorii obejmującej tego typu jednostki, co czyni zasób rosyjski przydatnym narzędziem w procesie poszukiwania reproduktów fatycznych i emotywnych.

Ostatnim z analizowanych w tym miejscu rosyjskich opracowań leksykograficznych jest *elektroniczna baza Pragmaticon*, pod którą należy rozumieć nowoczesne narzędzie językoznawcze służące opisowi tzw. formuł dyskursywnych (*дискурсивные формулы*). Niewątpliwie jest to projekt przełomowy na tle tradycyjnych słowników frazeologicznych i jednojęzycznych, stworzonych w ramach Narodowego Korpusu Języka Rosyjskiego. Baza ta zasługuje na szczególne uznanie nie tylko jako źródło danych językowych, lecz także jako modelowy przykład innowacyjnego podejścia do opisu wielowyrazowych jednostek języka mówionego, zwłaszcza tych o funkcjach reakcyjnych – fatycznych i emotywnych. Główna idea *Pragmaticonu* opiera się na traktowaniu wielowyrazowych formuł komunikacyjnych jako pełnoprawnych elementów systemu leksykalnego i zawiera między innymi takie konstrukcje, jak *а ты как думал; да ладно; не говори; ну а как же; вот именно*. Ich opis nie ogranicza się wyłącznie do warstwy semantycznej, lecz przede wszystkim uwzględnia funkcję pragmatyczną, sposoby użycia, warianty, gesty towarzyszące i konkretną sytuację komunikacyjną. Należy mocno podkreślić fakt, iż dane źródło elektroniczne rejestruje wyłącznie jednostki wielowyrazowe, co już samo w sobie czyni ją unikatowym zbiorem, zawierającym wielowyrazowce języka mówionego. Każda jednostka zawiera opis struktury wypowiedzi i ról rozmówców, dane o intonacji i prozodii (np. ИК-2), a także nagrania audio i wideo ukazujące rzeczywiste użycie formuły w kontekście sytuacyjnym. Uwzględnione zostają również tagi stylistyczne oraz semantyczne (np. *вводн., эмоц.*), określające funkcję i nacechowanie wyrażenia. Ujęcie takie znakomicie odwzorowuje pełen obraz języka potocznego, wykraczając poza tradycyjny model opracowania leksykograficznego. Charakterystyka formuł dyskursywnych opiera się na danych korpusowych, zaczerpniętych głównie z korpusu rosyjskiego oraz na założeniach lingwistyki kognitywnej i pragmatycznej. Szczególnie istotną cechą wyróżniającą dany słownik jest fakt, że niektóre formuły opatrzone są również tłumaczeniem na język angielski, co wyraźnie poszerza zakres funkcjonalny *Pragmaticonu* i sprawia, że projekt ten znacząco przekracza granice leksykografii jednojęzycznej, wchodząc w obszar słownictwa przekładowego i konfrontatywnego. Wskazane opracowanie wyznacza zatem nową tradycję i metodologię w opisie jednostek języka mówionego, stanowiąc narzędzie, które w znakomitym stopniu odpowiada potrzebom przeciętnych użytkowników, a nie tylko zaawansowanym językoznawcom. Jego dostępność, przejrzystość oraz nowoczesna forma elektroniczna czynią go przykładem leksykografii przyszłości, tj. rejestracji słownictwa, zakorzenionej w praktyce komunikacyjnej.

1.8 Rejestracja jednostek fatycznych i emotywnych we współczesnych czeskich słownikach jednojęzycznych

Współczesna czeska leksykografia jednojęzyczna rozwinęła się w XX wieku jako rezultat intensyfikacji badań językoznawczych, rosnącej potrzeby systematycznego opisu i uporządkowania materiału językowego. W danym okresie ukształtował się bardziej złożony układ typologiczny prac leksykograficznych: wraz z nowoczesnymi słownikami objaśniającymi rozwijały się słowniki specjalistyczne, np. diachroniczne, dialektalne, frazeologiczne, terminologiczne czy frekwencyjne. Dzięki takiemu zróżnicowaniu możliwe stało się skupienie na precyzyjnym, funkcjonalnym i reprezentatywnym opisie współczesnej czeszczyzny, bez nadmiaru informacji encyklopedycznych czy historycznych (Hladká 2005: 140–159). Do najważniejszych osiągnięć tego nurtu należy *Průruční slovník jazyka českého* (1935–1957), pierwszy pełny jednojęzyczny słownik objaśniający w językach słowiańskich, opracowany przez zespół czeskich językoznawców pod kierunkiem Oldřicha Hujera, Emila Smetánky i Miloše Weingarta (Hodura 1959: 6–8). Opracowanie to liczy około 250 tysięcy haseł, bazuje na szerokim materiale źródeł pisanych i cechuje się układem alfabetycznym oraz rozbudowaną mikrostrukturą, zawierającą, m.in.: dane gramatyczne, definicje, kwalifikatory stylistyczne i przykłady użycia. Na podkreślenie zasługuje rozwinięty repertuar kwalifikatorów (np. *hovor.*, *lid.*, *vulg.*), który pozwolił na precyzyjne zakwalifikowanie szeregu jednostek do konkretnych rejestrów. Mimo charakteru nienormatywnego, słownik ten odegrał istotną rolę w kodyfikacji czeskiego słownictwa oraz ugruntował wzorzec opisu języka ogólnego, uwzględniającego zróżnicowanie pragmatyczno-stylistyczne.

Równolegle do prac nad *PSJČ*, lecz niezależnie pod względem autorskim, koncepcyjnym i materiałowym, powstał *Slovník jazyka českého* autorstwa Františka Trávníčka, którego czwarte wydanie datuje się na rok 1952. Głównym celem danej pracy była rejestracja współczesnego zasobu leksykalnego czeszczyzny zarówno w wersji pisanej, jak i mówionej, z uwzględnieniem leksyki literackiej, regionalizmów oraz rozpowszechnionych zapożyczeń. Powstały w okresie intensywnych przemian społeczno-politycznych łączy rozbudowaną strukturę haseł, uwzględniającą dane gramatyczne, stylistyczne oraz pragmatyczne z materiałem ilustracyjnym odzwierciedlającym ideologiczny kontekst epoki. Z metodologicznego punktu widzenia stanowi przykład starannego i praktycznego podejścia do opisu języka narodowego, wyraźnie osadzonego w czeskiej tradycji filologicznej. W kontekście analizy jednostek wielowyrazowych, są one obecne przede wszystkim w postaci frazeologizmów, idiomów i konstrukcji składniowych, często umieszczanych w obrębie haseł

głównych jako ilustracje znaczeń lub przykłady użycia. Formy takie, jak *bojovati až do posledního dechu*; *opít se do němoty*; *ten se má!* czy *pracovati do západu slunce* wskazują, że autor dostrzegał ich wysoką frekwencję użycia w języku, co uzasadniało ich uwzględnienie w strukturze opisu leksykalnego. Jednostki o charakterze emotywnym reprezentowane są w słowniku przez zwroty nacechowane emocjonalnie, takie jak *ta radost doma!* czy *ta hanba!*, pełniące funkcję wykrzyknikową oraz ekspresywną. Choć struktury o funkcji fatycznej nie zostały graficznie wyodrębnione, a ich obecność w słowniku jest wysoce ograniczona, niektóre z nich można dostrzec w przykładach użycia. Można zatem podsumować, iż słownik Trávnicka ukazuje wielowymiarowy obraz czeskiej leksyki, uwzględniając elementy języka pisanego i mówionego, a także rejestrując utrwalone połączenia wyrazowe o różnorodnych funkcjach. Zaprezentowany w niniejszej pracy materiał pozwala wstępnie zarysować kierunek dalszych analiz nad jednostkami emotywnymi i fatycznymi oraz uchwycić ich funkcjonowanie w konkretnym kontekście sytuacyjnym.

Do osiągnięć czeskiej leksykografii XX wieku odnieść należy również pracę *Slovník spisovného jazyka českého*, opracowany przez zespół badaczy z Ústavu pro jazyk český Czeskiej Akademii Nauk pod kierownictwem Bohuslava Havránka. Słownik ten miał być kontynuacją opracowania *Příruční slovník jazyka českého*, jednak w przeciwieństwie do niego pełnił głównie funkcję normatywną, wyznaczając wzorzec współczesnej czeszczyzny literackiej opartej na uzusie pisanym. Celem autorów było stworzenie praktycno-dydaktycznego opracowania leksykograficznego, co zaowocowało przejrzystą strukturą haseł, rozbudowaną klasyfikacją stylistyczną i opisem ponad 190 tysięcy jednostek – od słownictwa ogólnego i specjalistycznego po formy potoczne czy regionalne. *SSJČ* cechuje się standaryzowaną strukturą artykułów hasłowych z informacjami gramatycznymi, kwalifikatorami stylistycznymi (np. *spisovné*, *hovorové*, *expr.*), definicjami oraz przykładami użycia, przy szczególnym uwzględnieniu podziału znaczeń i hierarchii polisemii. Dzięki wykorzystaniu szerokiej bazy materiałowej, zawierającej literaturę, prasę i dokumenty użytkowe, przedstawia on spójny i reprezentatywny obraz czeszczyzny literackiej XX wieku. Do momentu ukazania się specjalistycznych słowników idiomatycznych pełnił również funkcję pomocniczego zasobu frazeologicznego, kodyfikując liczne połączenia frazeologiczne i idiomatyczne w obrębie haseł głównych. Pomimo tej zalety jednostki wielowyrazowe o charakterze fatycznym i emotywnym praktycznie nie są w nim rejestrowane. Nieliczne wyrażenia ekspresywne (np. *ty jo!*; *no to snad ne!*) pojawiają się jedynie jako przykłady użycia lub cytaty, bez ich graficznego wyodrębniania i oddzielnej klasyfikacji. Nie odnotowano również typowych struktur fatycznych, takich jak formuły rozpoczynające i kończące

wypowiedź, co dodatkowo potwierdza ukierunkowanie słownika na język pisany, oficjalny i normatywnie uporządkowany. Należy wspomnieć, iż wersja elektroniczna słownika, dostępna od 2011 roku na stronie Instytutu Języka Czeskiego (ssjc.ujc.cas.cz), nie wprowadziła zasadniczych zmian w tym zakresie. Zawiera ona nie tylko pełną treść wydania drukowanego, lecz także materiały uzupełniające, jednak mimo rozszerzenia bazy nie rejestruje dodatkowych jednostek wielowyrazowych charakterystycznych dla języka mówionego. Obecne są natomiast liczne kolokacje i idiomy, jednak brak wśród nich interesujących nas replektów emotywnych i fatycznych. Podsumowując, dane opracowanie odegrało kluczową rolę w normatywizacji i opisie współczesnej czeszczyzny literackiej, stając się punktem odniesienia dla kolejnych słowników. Z perspektywy badań nad jednostkami wielowyrazowymi jego ograniczony zakres odzwierciedla dominującą wówczas tendencję autorów do koncentrowania się na opisie języka pisanego.

Za najbardziej reprezentatywne dzieło czeskiej leksykografii jednojęzycznej w zakresie systemowego opisu jednostek wielowyrazowych uchodzi pięciotomowy *Slovník české frazeologie a idiomatiky* pod redakcją Františka Čermáka. Słownik ten, publikowany w latach 1983–2009, wyróżnia się następującą strukturą:

- 1) *Přirovnání* – porównania i przysłowia;
- 2) *Výrazy neslovesné* – jednostki rzeczownikowe, przymiotnikowe, przysłówkowe oraz formy gramatyczne;
- 3) *Výrazy slovesné* – jednostki czasownikowe;
- 4) *Výrazy větné* – jednostki zdaniowe, w tym fatyzmy i emotywy;
- 5) *Onomaziologický slovník* – indeks znaczeniowy umożliwiający wyszukiwanie frazemów według ich funkcji semantycznej.

Na tle czeskiej, polskiej i rosyjskiej leksykografii jednojęzycznej opracowanie to stanowi jedyne tak pełne i kompleksowe ujęcie jednostek wielowyrazowych, uwzględniające zarówno tradycyjną frazeologię, jak i nowsze typy struktur idiomatycznych i pragmatycznych. Rejestruje ponad 35 000 czeskich frazemów zakorzenionych w uzusie, a jego celem była ich całościowa i systemowa charakterystyka z uwzględnieniem m.in. aspektów semantycznych, składniowych czy stylistycznych. Wszystkie tomy zostały opracowane według jednolitej i przejrzystej struktury hasłowej, obejmującej formę podstawową (wraz z możliwymi wariantami), charakterystykę składniową, definicję wzbogaconą o opis funkcjonalny oraz przykłady użycia zaczerpnięte z korpusu językowego. W kontekście badań nad jednostkami fatycznymi i emotywnymi szczególne znaczenie ma tom czwarty, rejestrujący najczęstsze frazemy języka mówionego. W ich składzie można odnaleźć formuły fatyczne (np. *Tak co?*;

Zatím se měj!), wyrażenia emotywne (np. *To je skvělý!*; *Neblázni!*) oraz wypowiedzi oceniające i ironiczne (np. *No to jsme to dopracovali*; *To je mi novinka*). Jednostki te opatrzone informacjami dotyczącymi ich funkcji (np. pożegnanie, wyraz zniecierpliwienia, komentarz ironiczny), stylu, typowych kontekstów użycia oraz wariantów, w tym także intonacyjnych i leksykalnych. Warto również zaznaczyć, że tom ten zawiera tłumaczenia wybranych jednostek na cztery języki obce (angielski, niemiecki, francuski i rosyjski), co znacząco podnosi jego przydatność w badaniach przekładowych. Ostatni piąty tom *Onomaziologický slovník*, opublikowany w 2009 roku, umożliwia wyszukiwanie jednostek według kategorii znaczeniowych, takich jak radość, gniew, zdziwienie, kontakt czy pochwała, i stanowi tym samym cenne źródło w analizach porównawczych oraz kontrastywnych. Z perspektywy badań nad frazemami oraz ich zastosowania w analizach kontrastywnych należy uznać, że dany słownik stanowi najbardziej kompleksowe i metodycznie spójne opracowanie jednostek wielowyrazowych w leksykografii czeskiej. Połączenie szczegółowego opisu jednostek z ich funkcjonalnym aspektem komunikacyjnym sprawia, że słownik ten poprzez swoją strukturę, zakres i metodologię reprezentuje wzorzec opisu frazematyki i może służyć jako punkt odniesienia zarówno dla dalszych prac jednojęzycznych, jak i przekładowych. Uzupełnienie wersji drukowanej o formę elektroniczną, wzbogaconą np. o nagrania intonacji, mogłoby dodatkowo zwiększyć jego użyteczność i zasięg badawczy.

Jako pierwsze z opracowań leksykograficznych XXI wieku w niniejszym opisie uwzględniono drugie, rozszerzone wydanie opracowania *Slovník nespisovné češtiny* pod redakcją Jana Huga, które stanowi jedno z bardziej rozbudowanych słowników poświęconych czeskiej leksyce niestandardowej. Drugie, rozszerzone wydanie słownika ukazało się w 2006 roku i liczy ponad 14 000 haseł, pochodzących z takich źródeł, jak literatura, internet, relacje ustne czy materiały zebrane przez użytkowników języka. Podstawowym celem słownika była dokumentacja czeszczyzny potocznej i slangowej, zwłaszcza tych obszarów, które nie posiadają swojego opisu w słownikach normatywnych. W słowniku dominują jednostki języka młodzieżowego, więziennego, wojskowego oraz gwar miejskich z Brna i Ostrawy. Pod względem mikrostruktury artykuły hasłowe opatrzone zostały kwalifikatorami stylistycznymi i środowiskowymi, często zawierają cytaty, informacje etymologiczne oraz przykłady użycia. Z perspektywy analizy jednostek wielowyrazowych dane opracowanie nie oferuje ich systematycznej rejestracji jako odrębnych haseł, a tego rodzaju jednostki występuje głównie w cytatach i przykładach użycia, pełniąc przy tym funkcje egzemplifikacyjne. Przykłady takie jak *potřeboval bych s tím helpnout*; *mluvit po našem* czy *snídaně na vidličku* świadczą o obecności ustabilizowanych kolokacji, często o charakterze regionalnym lub ekspresywnym. Pomimo

faktu, iż nie zostają one formalnie wyodrębnione jako frazeologizmy, ich kodyfikacja słownikowa potwierdza realne zakorzenienie tych połączeń w różnych aktach komunikacyjnych. W słowniku nie odnotowano również typowych jednostek o charakterze emotywnym czy fatycznym w formie samodzielnych artykułów hasłowych. W materiale ilustracyjnym można, co prawda, znaleźć wyrażenia nacechowane ekspresywnie (np. *magorák, hulibrk, švindlgrif*) lub zawierające elementy fatyczne (*pujdem na gáblik, ne?*), jednak ich obecność ma raczej charakter tekstowy niż systemowy. Podsumowując, *Slovník nespisovné češtiny* dostarcza bogatego materiału dokumentacyjnego dotyczącego leksyki potocznej i środowiskowej, w którym jednostki wielowyrazowe nie są rejestrowane w sposób całościowy, jako uzupełnienie haseł głównych. Analiza kontekstów użycia, w których występują może jednak przyczynić się do badań nad reproduktami języka mówionego oraz mechanizmami językowymi charakterystycznymi dla leksyki niestandardowej.

Do opracowań leksykografii jednojęzycznej XXI wieku odnieść należy również *Slovník současné češtiny*, opracowany przez zespół wydawnictwa Lingea, który ma za zadanie rejestrację współczesnej czeszczyzny z uwzględnieniem uzusu. Warto zaznaczyć, iż materiał leksykalny zawarty w słowniku został zgromadzony na podstawie autorskiego korpusu, co zapewniło reprezentatywność i aktualność opisywanych jednostek. Opracowania artykułów hasłowych cechują się przejrzystą i ustandaryzowaną strukturą, w ramach której autorzy wyróżniają, m.in. lemat, klasyfikację gramatyczną oraz znaczenia uporządkowane numerycznie. Integralną część każdego hasła stanowią liczne przykłady użycia oraz kwalifikatory stylistyczne (np. *hovor., expr., vulg., dětsky*) oraz środowiskowe (np. *slang.*), a także zwroty idiomatyczne, kolokacje czy frazeologizmy. Prezentowane jednostki wielowyrazowe uzupełniają strukturę artykułu hasłowego i nie funkcjonują jako odrębne artykuły hasłowe, co odzwierciedla utrwalony model stosowany w leksykografii jednojęzycznej. Jedną z istotnych zalet słownika jest bogata reprezentacja jednostek wielowyrazowych o charakterze idiomatycznym i frazeologicznym (np. *vzít roha; uložit si do paměti; být na větví*) oraz uwzględnienie typowych kolokacji, takich jak *držet hladovku; zacpat si uši*. Znacznie słabiej uwzględnione są jednostki fatyczne i ematywne. Słownik zawiera wprawdzie pojedyncze jednowyrazowe rejestracje tychże elementów (np. *jé; ty jo; ach; čau; nazdar*), jednak nie tworzą one samodzielnej kategorii i pojawiają się wybiórczo, najczęściej jako część przykładów użycia. Typowe reprodukt fatyczne i ematywne pozostają zatem poza polem widzenia analizowanego słownika, które koncentruje się przede wszystkim na czeskim słownictwie współczesnym, frazeologii oraz najbardziej znanych połączeniach idiomatycznych. Mimo otwarcia na leksykę potoczną i ekspresywną, interesujące nas

wielowyrzowce są w nim reprezentowane w sposób marginalny, co potwierdza szerszą tendencję do pomijania tego typu struktur w leksykografii jednojęzycznej.

Ostatnim z analizowanych czeskich słowników jednojęzycznych jest *Akademický slovník současné češtiny*, opracowywany od 2012 roku przez zespół pracowników Instytutu Języka Czeskiego Akademii Nauk Republiki Czeskiej. Celem tego opracowania jest deskryptywny opis współczesnej czeszczyzny, rejestrujący zarówno komunikację oficjalną, jak i codzienną od 1945 roku. Istotną zaletą słownika jest jego forma elektroniczna, umożliwiającą sukcesywne uzupełnianie bazy w porządku alfabetycznym, która docelowo ma objąć od 120 000 do 150 000 jednostek leksykalnych. Nawiązuje on również do tradycji czeskich słowników jednojęzycznych, takich jak wspomniane *PSJČ* i *SSJČ*, wprowadzając jednocześnie nowe rozwiązania metodologiczne, m.in. korpusowe podstawy opisu, rozszerzoną analizę gramatyczną oraz systematyczną klasyfikację stylistyczną i frazeologiczną. Struktura artykułu hasłowego jest przejrzysta i rozbudowana oraz uwzględnia takie informacje, jak lemat, wymowę, odmianę, etymologię, kwalifikatory stylistyczne i tematyczne, definicje, przykłady użycia, frazeologizmy oraz odsyłacze do haseł pokrewnych i gniazdowanych. Możliwe jest również filtrowanie wyszukiwań według konkretnych kategorii gramatycznych (np. czasowniki, wykrzykniki, frazy), co sprawia, że słownik ma wartość praktyczną zarówno dla szerokiego grona użytkowników, jak i dla badaczy zajmujących się szczegółową analizą języka. Szczególną wartość słownika w badaniach nad jednostkami wielowyrzowymi, w tym fatycznymi i emotywnymi, stanowi możliwość ich wyszukiwania zarówno w postaci samodzielnych haseł, takich jak *čertví co*; *čert ví co*; *co si to dovoluješ* czy *co to dá*, jak i jako elementów odsyłających do haseł głównych (np. *co čert nechťel* lub *ale co* i *ale což*). Jednostki te zostają opatrzone odpowiednimi kwalifikatorami, takimi jak *kolokviální* czy *expresivní*, mają przypisane synonimy, np. *čert ví co ~ kdovíco, bůhví co*, oraz są ilustrowane przykładami użycia zaczerpniętymi z rzeczywistego użycia w sytuacji komunikacyjnej. Można zaobserwować również rozszerzony repertuar czeskich frazeologizmów, przysłów i idiomów, takich jak *číst někomu myšlenky*; *zachovat chladnou hlavu*; *někdo si myslí, že sežral Šalamounovo hovno* czy *veselá mysl je půl zdraví*. Jednostki te są systematycznie klasyfikowane, opisywane i ilustrowane, co czyni słownik cennym źródłem dla badań nad pragmatyką i semantyką tekstu. Na podkreślenie zasługuje również zakładka *Sousloví a frazémy*, dzięki której możliwe jest odnalezienie licznych struktur niebędących typowymi kolokacjami czy związkami frazeologicznymi, lecz realizujących konkretne funkcje pragmatyczne w wypowiedzi. Można zatem stwierdzić, iż *Akademický slovník současné češtiny* to nowoczesne narzędzie leksykograficzne, czerpiące z korpusów jednojęzycznych i rozwiązań cyfrowych, które

systematycznie uwzględnia nie tylko ogólny zasób współczesnej czeszczyzny. Poświęcenie osobnego opisu jednostkom wielowyrazowym o funkcji fatycznej i ematywnej świadczy o kontynuacji tej tendencji w leksykografii czeskiej, polegającej na uwzględnianiu nie tylko materiału języka pisanego, lecz także jednostek charakterystycznych dla języka mówionego. Systematyczne rozszerzanie słownika o nowe jednostki leksykalne i kolejne sekcje alfabetyczne niewątpliwie zwiększy jego wartość jako źródła do analizy jednostek wielowyrazowych, zwłaszcza w kontekście ich struktury, wariantów i funkcji w wypowiedzi.

1.9 Leksykografia przekładowa wobec wielowyrazowych jednostek języka mówionego

Współczesna leksykografia przekładowa, pojmowana jako dziedzina zajmująca się opracowywaniem słowników dwu- i wielojęzycznych oraz ich teoretycznym opisem, pozostaje ściśle związana z teorią i praktyką przekładu. Jej celem jest nie tylko rejestrowanie jednostek leksykalnych danego języka, lecz także umożliwienie ich przekładu na język obcy z uwzględnieniem znaczenia, funkcji komunikacyjnej oraz nacechowania stylistyczno-pragmatycznego. Jak zauważa Ju. E. Dorochowa, teoria przekładu od początku rozwijała się w kontekście interdyscyplinarnym, czerpiąc z dorobku różnych dziedzin, a szczególnie silne związki łączą ją z leksykografią, która dostarcza tłumaczowi wiedzy słownikowej, kulturowej i specjalistycznej, niezbędnej do skutecznego przekładu. Autorka wskazuje ponadto, że sama praktyka przekładu stanowi istotne źródło dla rozwoju leksykografii przekładowej:

«[...] сама практика перевода является основой для создания и пополнения двуязычных словарей. Именно на основе многовековой практики межъязыковых контактов сложился переводческий опыт и выработалась лексикографическая традиция» (Дорохова 2013: 74). Na potrzebę uwzględnienia wymiaru funkcjonalno-stylistycznego zwraca uwagę również Bogumił Gasek, który podkreśla, iż celem leksykografii przekładowej jest nie tylko wyjaśnienie znaczenia jednostki leksykalnej języka wyjściowego, ale także dobór jej funkcjonalnego i stylistycznego odpowiednika w języku docelowym (Gasek 2020: 90). Skuteczne tłumaczenie wymaga nie tylko znajomości systemów leksykalnych obu języków, lecz także umiejętności dostrzegania różnic pragmatycznych i kulturowych, ujawniających się zwłaszcza w jednostkach funkcjonujących poza rejestrem oficjalnym – zależnych od kontekstu sytuacyjnego, relacji społecznych i intencji mówiącego. Ich przekład okazuje się szczególnie wymagający, ponieważ wiele z tych wyrażen nie znajduje kodyfikacji w opracowaniach leksykograficznych z powodu swojej idiomatyczności, ograniczonej frekwencji, nacechowania

stylistycznego lub braku jednoznacznych odpowiedników. Podobne obserwacje formułują również czescy badacze, którzy zauważają, że „při porovnávání zpracovávaných jazyků si slovníkár musí stále udržovat patřičnou distanci od vlivu jak cizího jazyka, tak i své mateřštiny. Jelikož je asymetrie významu běžným jevem, ozřejmuje se téměř na každém slově složitost ekvivalentačního procesu” (Vachková 2019: 104).

W obszarze leksykografii przekładowej, gdzie głównym przedmiotem zainteresowania są zazwyczaj ustabilizowane jednostki systemowe, szczególnym wyzwaniem pozostaje rejestracja i opis jednostek wielowyrazowych, zwłaszcza tych charakterystycznych dla języka mówionego, nacechowanych emocjonalnie, fatycznie, potocznie lub sytuacyjnie. Analiza słowników wielojęzycznych pokazuje, że konstrukcje potoczne, idiomatyczne i silnie zależne od kontekstu rzadko są rejestrowane ze względu na wieloznaczność, rozmyte granice znaczeniowe i możliwość istnienia wielu ekwiwalentów przekładowych. Pomimo rosnącej świadomości potrzeby rejestrowania jednostek wielowyrazowych typowych dla języka mówionego, słowniki przekładowe wciąż koncentrują się głównie na leksyce ogólnej, przez co rzadko oddają pełen obraz funkcjonowania oraz przekładalności bardziej złożonych konstrukcji.

Jak zauważają A. I. Małkowa i L. N. Mirosznichenko, słowniki przekładowe wciąż nie są wystarczająco wyposażone w narzędzia umożliwiające pełne odwzorowanie stylistycznej i funkcjonalnej różnorodności jednostek złożonych (Малкова, Мирошниченко 2019: 422). Tymczasem to właśnie język mówiony, nacechowany emocjonalnie, fatycznie lub sytuacyjnie, generuje najwięcej trudnych do przetłumaczenia, a zarazem najczęściej używanych konstrukcji. Praktyka przekładowa pokazuje jednak, że nawet konstrukcje uznawane za częste bywają w słownikach pomijane, co rodzi potrzebę ich współczesnej systematycznej kodyfikacji i tłumaczenia na języki obce.

W dotychczasowych opracowaniach leksykografii przekładowej jednostki wielowyrazowe były najczęściej ujmowane jako rozwinięcia haseł głównych, w formie typowych kolokacji i idiomów wraz z ich przekładami na języki docelowe, natomiast rzadko występowały jako samodzielne jednostki hasłowe. Jedną z przyczyn takiego podejścia może być występowanie zjawiska poliekwiwalentności, rozumianego jako obecność wielu możliwych odpowiedników przekładowych dla jednej jednostki języka źródłowego. Jak zauważa W. N. Komissarow, «единичное соответствие означает, что в большинстве случаев данная единица ИЯ (исходный язык – А.З) переводится одной и той же единицей ПЯ (переводной язык – А.З)». Natomiast odpowiedniki wielorakie to «несколько единиц ПЯ, регулярно используемых для передачи их значений» (Комиссаров 2011: 167). To

właśnie ta druga kategoria wiąże się bezpośrednio ze zjawiskiem poliekwiwalentności (nazywanym również wieloeckiwalentnością), które w literaturze przedmiotu opisywane jest także za pomocą terminów takich jak koekiwalenty, kowarianty czy dubletowe odpowiedniki przekładowe (Дзене, Перевышина 2013: 13). W takich przypadkach leksykograf powinien nie tyle dążyć do wskazania jednego, uniwersalnego ekwiwalentu, ile uwzględniać zmienność wariantów oraz ich uwarunkowania pragmatyczne, które w sposób istotny wpływają na wybór tłumaczenia. Z tego względu szczególną rolę odgrywa kompetencja tłumacza, zobowiązanego do uwzględniania wszystkich aspektów znaczeniowych jednostki w określonym kontekście sytuacyjnym.

G. W. Kołszanskij wyróżnia w tym zakresie dwa zasadnicze typy kontekstu: przedmiotowy, odnoszący się do kategorii takich jak czas, miejsce czy relacje pomiędzy przedmiotami, oraz sytuacyjny, uwzględniający warunki i przebieg aktu komunikacji (Колшанский 1980: 32–39). Skuteczne tłumaczenie wymaga zatem od tłumacza samodzielnego rozszerzania kontekstu użycia wyrażeń przy wykorzystaniu zarówno wiedzy językowej, jak i pozajęzykowej. Jak w rozumieniu przytoczonego językoznawcy, kontekst nie stanowi jedynie jednego z elementów procesu translacyjnego, lecz warunkuje jego twórczy wymiar jako aktu interpretacji tekstu w jego aspektach kulturowym, poznawczym i estetycznym.

W świetle powyższych rozważań coraz wyraźniej ujawnia się potrzeba opracowania podejścia leksykograficznego, które uwzględniałoby realne, uzualne użycie języka, a nie jedynie jego znormalizowaną, systemową postać. Szczególną uwagę należy w nim poświęcić systematycznej rejestracji jednostek wielowyrazowych, będących istotnym składnikiem praktyki językowej, a mimo to rzadko ujmowanych jako odrębne hasła. Tymczasem to właśnie konstrukcje nacechowane emocjonalnie, fatycznie i sytuacyjnie odgrywają istotną rolę w komunikacji i wymagają ujęcia nie tylko technicznego, lecz także interpretacyjnego. W dalszej części pracy przedstawiona zostanie analiza wybranych słowników dwujęzycznych: polsko-rosyjskich, rosyjsko-polskich, polsko-czeskich i czesko-polskich, ze szczególnym uwzględnieniem obecności jednostek wielowyrazowych, zakresu ich rejestracji i reprezentacji funkcjonalno-pragmatycznej.

1.10 Rejestracja jednostek fatycznych i emotywnych we współczesnych słownikach polsko-rosyjskich i rosyjsko-polskich

Do najważniejszych słowników przekładowych polsko-rosyjskich XX wieku zaliczyć należy *Большой польско-русский словарь* autorstwa Dymitra Hessena i Ryszarda Stypuły, który był wielokrotnie wznawiany od 1955 roku. Opracowanie to powstało z myślą o szerokim gronie użytkowników, w tym tłumaczach, nauczycielach, studentach, i stanowi rozwinięcie inicjatyw leksykograficznych podejmowanych w związku z intensyfikacją kontaktów polsko-rosyjskich. Słownik liczy około 80 000 jednostek, w tym leksykę ogólną i specjalistyczną oraz liczne konstrukcje wielowyrazowe (np. frazeologizmy, idiomy, kolokacje), a także wyrażenia metaforyczne czy archaiczne – zarówno systemowe, jak i posiadające konkretne cechy pragmatyczne. Znaczący odsetek haseł to konstrukcje nacechowane stylistycznie, takie jak wyrażenia potoczne, żartobliwe i wulgarne, a ich budowa uwzględnia, m.in.: formy fleksyjne, przykłady użycia czy różne typy tłumaczeń, od ekwiwalentów bezpośrednich po parafrazy kontekstowe w języku rosyjskim. Większość jednostek frazeologicznych stanowi uzupełnienie obszerniejszych artykułów hasłowych i tylko wybiórczo zostają typograficznie wyróżnione za pomocą znaków takich jak ♦ czy ~, przy czym rzadko występują one jako odrębne hasła.

Z perspektywy badań nad wielowyrazowymi ematywami i fatyzmami szczególnie znaczenie mają zarejestrowane w słowniku wyrażenia, takie jak *a to dopiero numer!* (*вот так номер!*); *co za czasy!* (*что за времена!*); *nie do wiary!* (*невероятно!*), stanowiące przydatny materiał do weryfikacji polsko-rosyjskich par przekładowych. Obok nich pojawiają się także krótkie wyrażenia o charakterze fatycznym, m.in.: *jak się masz?* (*как поживаешь?*); *halo!* (*алло!*); *co słyhać?* (*как дела?*), pełniące funkcję nawiązywania lub podtrzymywania kontaktu. Szczególnie istotna pozostaje świadoma kodyfikacja reproduktów, łączących funkcje ematywne i fatyczne, takie jak *a nie mówiłem?* (*разве я не говорил?*) czy *a ładnie to tak?* (*разве так можно?*), które, choć nie posiadają samodzielnych haseł, są tłumaczone z uwzględnieniem ich intencji komunikacyjnej. Można podsumować, że opracowanie zawiera rozbudowany zasób formuł i jednostek wielowyrazowych (np. *być w złym humorze*; *między młotem a kowadłem*; *wyprowadzić z równowagi*), mimo że nie zostały one wyodrębnione jako oddzielna kategoria w makrostrukturze słownika. Dzięki połączeniu ekwiwalencji leksykalnej ze świadomym doбором odpowiedników słownik zachowuje wysoką wartość praktyczną i naukową, jednocześnie wyznaczając kierunek rozwoju leksykografii przekładowej, w którym jednostki ematywne i fatyczne traktowane są jako integralne elementy opisu.

Jednym z najobszerniejszych dwujęzycznych i dwukierunkowych opracowań leksykograficznych XXI wieku jest *Nowy słownik rosyjsko-polski / polsko-rosyjski* autorstwa Jana Wawrzyńczyka z 2008 roku. Słownik liczy około 23 000 haseł w każdej z części i wyróżnia się nowoczesną, przejrzystą strukturą, opartą na uproszczonej transkrypcji, znaczeniowo uporządkowanych odpowiednikach przekładowych oraz licznych kolokacjach. Struktura artykułu hasłowego uwzględnia między innymi informacje morfologiczne oraz komentarze stylistyczne, a zastosowanie intuicyjnych symboli graficznych zamiast wcześniej stosowanych, mniej czytelnych znaków (takich jak tylda czy nawiasy ostrokątne), ułatwia korzystanie z opracowania zarówno osobom uczącym się języka, jak i tłumaczom. Materiał leksykalny skupia się przede wszystkim na języku ogólnym i potocznym, pomijając formy substandardowe i ekspresywne, dzięki czemu słownik znajduje zastosowanie w dydaktyce i przekładzie, choć w mniejszym stopniu nadaje się do analizy języka mówionego czy potocznego. Z perspektywy badań nad wielowyrazowymi jednostkami fatycznymi i emotywnymi, owa praca okazuje się jednak narzędziem o ograniczonej przydatności. Pomimo rejestracji licznych kolokacji czy idiomów, nie oferuje on rozbudowanej warstwy jednostek języka mówionego, używanych w komunikacji codziennej. Egzemplifikacyjnie ilustruje to hasło *jak*, w którym obok podstawowych znaczeń umieszczono m.in. *jak się masz?* (*как поживаешь?*); *jak to?* (*как это?*); *jakby co (to...)?* (*если что – то...*); *jak Boga kocham* (*клянусь Богом*) czy *jakby nigdy nic* (*как ни в чём не бывало*), co potwierdza rejestrację najczęściej używanych konstrukcji, bez ich systematycznej klasyfikacji funkcjonalnej. Kolejnym przykładem może być hasło *porządek*, w którym pojawiają się połączenia leksykalne typu *w porządku* (*в порядке*); *w porządku rzeczy* (*в порядке вещей*); *w porządku przyjaźni* (*на основе дружбы*) czy *w kolejności / według kolejności* (*в порядке / по очереди*). Choć niektóre z tych jednostek mogą być używane jako reakcje na wypowiedź nadawcy i pełnić funkcję potwierdzająco-fatyczną, trudno traktować je jako systemowo opracowane przykłady produktów fatycznych. Należy podkreślić, że pomimo wyraźnie ograniczonej rejestracji jednostek fatycznych i emotywnych analizowany słownik stanowi starannie opracowane narzędzie translatorskie, wyróżniające się bogatym zasobem leksykalnym oraz funkcjonalnymi rozwiązaniami redakcyjnymi. Ze względu na brak wyodrębnienia i klasyfikacji funkcji pragmatycznych, opracowanie nie znajduje pełnego zastosowania w badaniach nad jednostkami fatycznymi i emotywnymi, jednak pozostaje cennym źródłem uzupełniającym w analizach kontrastywnych, translatorycznych oraz w dydaktyce.

Kolejnym z analizowanych opracowań jest *Polsko-rosyjski słownik frazeologiczny gwary młodzieżowej* autorstwa Eleny Nevzorovej-Kmech z 2018 roku, ukierunkowany na

rejestrację par przekładowych charakterystycznych dla młodzieżowego języka potocznego. Publikację tę można sytuować w obszarze frazeologii stosowanej, badań nad socjolektami oraz leksykografii porównawczej przy szczególnym uwzględnieniu wyrażen nacechowanych emocjonalnie. Zasadniczym celem słownika było uzupełnienie luki wynikającej z braku polsko-rosyjskich opracowań tej odmiany języka oraz zaprezentowanie strategii przekładu wybranych jednostek z perspektywy dydaktycznej i translatorycznej. Materiał leksykalny stanowi kilkaset frazeologizmów z różnych rejestrów, takich jak gwara studencka, młodzieżowa, przestępcza czy muzyczna, którym można przypisać nacechowanie ironiczne, wulgarne lub groteskowe. Artykuły hasłowe mają układ alfabetyczny i zawierają kwalifikatory stylistyczne (np. *pot.*, *żart.*, *wulg.*, *eufem.*, *lekcew.*), komentarze etymologiczne, transformacyjne lub tłumaczenia dosłowne oraz przybliżone odpowiedniki slangowe, a niekiedy także parafrazy i odsyłacze. Szczególną wartość słownika stanowi znaczny zbiór jednostek emotywnych, takich jak frazy wykrzyknikowe (np. *jak babcie Kocham; jak bich mech; jak bonie dydy*) czy wyrażenia fatyczne (*jak tam gryzie?; jak żyjesz?*), traktowane jako samodzielne hasła. Istotną grupę tworzą również emotywy reakcyjne, takie jak *jak tu stoję; jak trza* czy *jak w banku*, pełniące funkcję ekspresywnych potwierdzeń lub zapewnień w odpowiedzi na zachowanie rozmówcy. Zdecydowana większość ekspresywnych jednostek wielowyrazowych cechuje się niejednoznacznością strukturą składniową i często stanowi humorystyczne przekształcenia znanych frazeologizmów (np. *jak Bożę Kocham/jak bułkę Kocham*), przy czym podkreślona zostaje ich funkcja wzmacniająca. Słownik wyróżnia się również rozbudowanymi kategoriami klasyfikacji, uwzględniającymi, m.in. środowisko użytkownika (np. *skate*, *muz.*, *przest.*), przy czym niemal całkowicie pomija jednostki fatyczne, koncentrując się głównie na funkcjach ekspresywnych, ironicznych i emocjonalnych. Pomimo faktu, iż wiele wyrażen zawartych w słowniku jest dziś rzadziej używanych i pełni głównie funkcje stylizacyjne, pozostaje on przydatnym źródłem do analizy wielowyrazowych emotywow oraz ich przekładu na język rosyjski.

Zupełnie nowy etap w leksykografii polsko-rosyjskiej zdaje się otwierać wydany w 2003 roku słownik *Idiomy polsko-rosyjskie*, opracowany przez Wojciecha Chlebdę, Annę Gołubiewą, Jana Wawrzyńczyka i Tomasza Wielga – specjalizujących się w leksykografii przekładowej, frazeologii i pragmatyce językowej. Opracowanie to jest jednym z pierwszych dwujęzycznych słowników poświęconych jednostkom frazeologicznym, które choć formalnie określone jako idiomy, wyraźnie wykraczają poza ramy klasycznie definiowanej idiomatyczności. Autorzy zdecydowali się zarejestrować szereg połączeń o charakterze potocznym, pragmatycznym, nierzadko pozbawionych idiomatycznego znaczenia, takich jak:

aparat nadawczy – передатчик; łyżka do butów – ложок; ani tu się śni – и не думает czy chyba że – разве что. Dobór materiału wskazuje na odejście od klasycznej frazeologii i czyni słownik punktem wyjścia do dalszych rozważań, w których centrum opisu leksykalnego stopniowo przesuwają się z tradycyjnych jednostek frazeologicznych na jednostki odtwarzalne, o wysokiej frekwencji i wielowyrazowe. Struktura haseł obejmuje układ dwujęzyczny z przykładami kontekstowymi, jednak w większości przypadków nie zawierają one kwalifikatorów stylistycznych. Zostało to skrytykowane przez recenzentów (Szczerbowski 2003: 379), ponieważ utrudnia użytkownikom prawidłowe rozpoznanie rejestru językowego tłumaczonych jednostek. Z perspektywy badań nad jednostkami fatycznymi i emotywnymi słownik ten jest pierwszym opracowaniem, w którym takie wyrażenia zostały systemowo wydzielone jako odrębne hasła i występują w stosunkowo dużej liczbie. Zawiera on liczne przykłady konstrukcji o charakterze emotywnym czy fatycznym, takich jak *a to dopiero! – вот ещё!; ani mi się śni! – и не подумаю!; jak się masz? – как поживаешь?; co słyszeć? – как дела?; a nie mówiłem? – разве я не говорил? czy a ładnie to tak? – разве так можно?*. Należy podkreślić, że słownik ten stanowi ważne osiągnięcie w leksykografii dwujęzycznej, które mimo pewnych braków w zakresie systematyki stylistycznej i pragmatycznej, wyróżnia się spójną koncepcją, innowacyjnym podejściem do idiomatyczności oraz poszerzeniem granic tradycyjnej frazeologii. Jego wartość zawiera się głównie w wyznaczeniu nowego kierunku w leksykografii przekładowej i otwarciu przestrzeni dla dalszych, pogłębionych analiz jednostek wielowyrazowych o funkcjach komunikacyjnych. To właśnie to opracowanie można uznać za załączek przyszłych idiomatykonów, rozwijających zapoczątkowane tu badania w kierunku kodyfikacji i przekładu reproduktów.

Pracą najistotniejszą z punktu widzenia podejmowanych w niniejszej rozprawie badań jest *Podręczny idiomatykon polsko-rosyjski* (2006–2019), opracowany przez zespół z Uniwersytetu Opolskiego pod kierunkiem Wojciecha Chlebdy. Stanowi on nowatorskie przedsięwzięcie słownikarskie, które zarówno pod względem doboru materiału, jak i sposobu jego prezentacji, wykracza poza tradycyjne ramy dotychczasowego opisu leksykograficznego. Jako idiomatykon –słownik tematyczny, komentowany, dydaktyczny – rejestruje wyłącznie jednostki wielowyrazowe, występujące w rzeczywistej komunikacji, którym można przypisać wysoką odtwarzalność w określonych sytuacjach, przy czym jednostki te nazywane są właśnie reproduktami. Potrzeba stworzenia tego słownika wynikała z celów dydaktycznych oraz z praktyki translacyjnej, ponieważ idiomatykon ma służyć studentom, tłumaczom i wszystkim uczącym się języka rosyjskiego jako narzędzie wspomagające przekład związków składniowych i fraz, które nie mieszczą się w tradycyjnych kategoriach idiomatycznych.

Materiał oparto o autentyczne obserwacje językowe, a całość została wydana w formie dziesięciu zeszytów, zebrane następnie w dwóch tomach zbiorczych z hasłami uporządkowanymi alfabetycznie oraz tabelarycznie. Zestawiane w słowniku polskie frazy z rosyjskimi odpowiednikami umożliwiają ich bezpośrednie wykorzystanie w przekładzie, bez potrzeby głębokiej rekonstrukcji składniowej. W przypadku różnic strukturalnych stosowana jest tzw. akomodacja składniowa. Idiomatykon nie podaje natomiast przykładów użycia, ale koncentruje się na wskazywaniu możliwych odpowiedników i diagnozowaniu luk ekwiwalencyjnych, co stanowi dodatkowy walor w badaniach kontrastywnych. Wyróżniającą cechą słownika jest również rozwinięta metodologia budowy artykułów hasłowych. Jednostki zapisywane są często w formie zdań w trzeciej osobie liczby pojedynczej (np. *kto nie kiwnął <nawet> palcem*) i oznaczane według ścisłych zasad: komponenty stałe pogrubieniem, zmienne kursywą, wymienne ukośnikiem prawym, a miejsca walencyjne kreską. Dla haseł czasownikowych stosuje się wyrazy naprowadzające (np. *łamać: kto łamie przepisy*). Po stronie rosyjskiej pojawiają się zarówno odpowiedniki bezpośrednie, jak i symbol Ø, wskazujący na brak systemowego ekwiwalentu i konieczność zastosowania innego rozwiązania translatorskiego. Szczególnie istotna z perspektywy prowadzonych badań jest rejestracja formuł fatycznych i emotywnych, którym poświęcono w słowniku osobne rozdziały, a ich łączną liczbę można szacować na około 300 jednostek. Wśród zarejestrowanych jednostek znajdują się m.in. formuły fatyczne, takie jak *No i co dalej?* (ros. *Ну и что дальше?*) czy *A kto by tam słuchał!* (ros. *Да кто бы его слушал!*), a także emotywy języka mówionego, np. *Serce mi pęka* (ros. *У меня сердце разрывается*) lub *No trzymaj mnie, bo nie wytrzymam!* (ros. *Держи меня, а то сорвусь!*). Podsumowując, *Podręczny idiomatykon polsko-rosyjski* stanowi przełomowe osiągnięcie w obszarze leksykografii przekładowej, będąc pierwszym tego typu opracowaniem skoncentrowanym wyłącznie na jednostkach wielowyrazowych w relacji polsko-rosyjskiej. Praca ta niewątpliwie redefiniuje tradycyjne podejście do makro- i mikrostruktury słownika, odchodząc od klasycznego modelu leksykografii idiomatycznej na rzecz rejestracji szerokiego spektrum reпродуктów – od idiomów i kolokacji po formuły pragmatyczne i tekstotwórcze – wywodzących się z niemal wszystkich sfer ludzkiej aktywności. Praca ta wyznacza nowy kierunek w opisie leksykograficznym zarówno pod względem metodologicznym, jak i koncepcyjnym. Jako dzieło zamknięte w dziesięciu zeszytach i dwóch tomach zbiorczych, stanowi imponujący rezultat ponad dziesięcioletniej pracy zespołu autorów, jednak jego potencjał nie został wyczerpany. Ze względu na skalę materiału oraz rozwijający się paradygmat frazeologii nadawcy, kontynuacja tego projektu wydaje się nie tylko uzasadniona, ale konieczna. Przedstawiona rozprawa doktorska stanowi

próbę rozwinięcia i dopełnienia tej idei poprzez skoncentrowanie się na jednej z najistotniejszych, a dotąd nie w pełni zbadanych kategorii reproduktów – jednostkach fatycznych i emotywnych. Celem jest nie tylko wyodrębnienie zasobu tych formuł w języku polskim, ale także ustalenie ich translacyjnych odpowiedników w języku rosyjskim i czeskim, co wpisuje się zarówno w tradycję idiomatykonu, jak i w szerszy nurt nowoczesnej leksykografii przekładowej.

1.11 Rejestracja jednostek fatycznych i emotywnych we współczesnych słownikach polsko-czeskich i czesko-polskich

Jednym z pierwszych opracowań leksykografii polsko-czeskiej XX wieku jest *Słownik polsko-czeski* autorstwa Bohumila Vydry, który w swoim trzecim wydaniu z 1967 roku stanowi klasyczne narzędzie leksykograficzne, opracowane z myślą o użytkownikach czeskich uczących się języka polskiego. Publikacja zawiera około 35 000 haseł i została opracowana jako praktyczne opracowanie dla celów dydaktycznych, edukacyjnych, kulturowych oraz technicznych. Wydanie to uznaje się za istotnie rozszerzone, uwzględniające nowe tendencje rozwojowe w polszczyźnie, zwłaszcza w zakresie słownictwa potocznego. Makrostruktura słownika opiera się na alfabetycznym układzie polskich haseł z uwzględnieniem formy podstawowej każdej jednostki oraz informacją o rodzaju, aspekcie, odmianie i komentarzem fleksyjnym. Z kolei mikrostruktura haseł zawiera przekłady na język czeski (często w kilku wariantach), kwalifikatory gramatyczne, oznaczenia aspektu, rekcję oraz liczne przykłady użycia w kontekście. Autor uwzględnił również typowe jednostki frazeologiczne i idiomatyczne, rejestrowane zarówno w obrębie haseł podstawowych, jak i jako odrębne wyrażenia, często oznaczane znakiem graficznym (♦). Przykłady takie jak *spadł mi ~ z serca* – *spadł mi kámen ze srdce*; *nie pozostawić ~ na kamieniu* – *nezůstal kámen na kameni*, jak *dwie krople wody* – *jako vejce vejci* czy *nic takiego nie widziałem* – *něco takového jsem neviděl* świadczą o ich istotnej roli w strukturze słownika.

Jednostki fatyczne i ematywne nie są wyodrębnione systemowo ani ujęte jako samodzielne hasła, jednak pojawiają się wrywkowo w obrębie frazeologii i idiomów, głównie jako formuły reakcyjne lub kontaktowe. Przykłady takie jak *jak się panu ~ba* – *jak si přejete* wskazują na ich obecność, choć ograniczoną do najczęściej używanych zwrotów. Można stwierdzić, iż na tle dwudziestowiecznych opracowań dwujęzycznych słownik Vydry wyróżnia się starannością redakcyjną, rozbudowaną bazą leksykalną oraz konsekwentnym uwzględnianiem jednostek wielowyrazowych. Pomimo stosowania tradycyjnych metod

kartotekowych, słownik ten pozostaje wartościowym źródłem do celów dydaktycznych i przekładoznawczych. Nie dostarcza jednak wystarczającego materiału do pogłębionych badań nad reproduktami fatycznymi i emotywnymi, pojawiającymi się w nim jedynie sporadycznie i bez ujęcia systemowego.

Z kolei do najobszerniejszych opracowań w obszarze czesko-polskiej leksykografii przekładowej końca XX wieku jest *Słownik czesko-polski* pod redakcją Janusza Siatkowskiego i Mieczysława Basaja z 1991 roku. Całość słownika liczy około 53 000 jednostek, z czego 28 000 haseł stanowią wyrażenia wielowyrazowe, przede wszystkim kolokacje, idiomy oraz frazeologizmy. Opracowanie to stanowi użyteczne narzędzie w badaniach nad ustabilizowanymi jednostkami leksykalnymi i frazeologicznymi języka czeskiego, zwłaszcza z perspektywy przekładu i dydaktyki. Mikrostruktura haseł została opracowana w sposób przejrzysty i uporządkowany według czeskiego alfabetu. W obrębie każdego hasła został uwzględniony podział na znaczenia oznaczone numeracją oraz konsekwentnie zastosowano kwalifikatory stylistyczne, takie jak *pot.* (potoczne), *expr.* (ekspresywne) czy *zgr.* (zgrubienia). Każdy z artykułów hasłowych zawiera również liczne przykłady użycia poprzez przytoczenie idiomów i fraz wyróżnionych graficznie, uzupełnionych o zwięzłe objaśnienia oraz informacje o aspekcie czasownikowym. Niniejsze rozwiązanie pozwala nie tylko na odnalezienie właściwego odpowiednika przekładowego, lecz także na zapoznanie się z funkcją oraz zakresem użycia danej jednostki w szerszym kontekście tekstowym. Wśród licznych jednostek wielowyrazowych znaleźć można zarówno ustabilizowane frazeologizmy (np. *házet klacky pod nohy* – *rzucać komuś kłody pod nogi*), jak i częste połączenia leksykalne oraz zwroty (np. *dobrou noc* – *dobranoc*; *držet dietu* – *trzymać dietę*). W słowniku znaleźć można pojedyncze, najczęściej używane formuły o charakterze fatycznym i emotywnym, takie jak *nazdar!* – *cześć!*; *tak co?* – *no i co?*; *jak se vede?* – *jak leci?*; *to je blbost!* – *to głupota!*; *je mi zle* – *jest mi źle*, jednak nie stanowią one osobnych haseł. W tym przypadku rejestracja reproduktów fatycznych czy emotywnych ma charakter ograniczony i uwzględnia głównie najczęstsze zwroty języka mówionego. Warto podkreślić, że słownik ma charakter konkatencyjny, tj. dla każdego znaczenia czeskiego leksemu podano jedno lub kilka najbliższych odpowiedników polskich, często wzbogaconych o komentarze semantyczne lub stylistyczne. Słownik Siatkowskiego i Basaja można zatem uznać za istotne narzędzie o wysokiej wartości dydaktycznej i przekładoznawczej, szczególnie użyteczne w pracy z językiem ogólnym, jednostkami idiomatycznymi, kolokacjami oraz utrwalonymi zwrotami współczesnej czeszczyzny. Nie jest to źródło w pełni reprezentatywne dla badań nad reproduktami fatycznymi i emotywnymi,

jednak może stanowić punkt wyjścia do dalszych analiz i poszukiwań jednostek wielowyrazowych wraz z ich potencjalnymi odpowiednikami przekładowymi.

Na początku XXI wieku ukazało się obszerne, dwutomowe opracowanie *Polsko-český slovník* pod redakcją Karla Olivy, którego pierwszy tom opublikowano w 2006 roku, a drugi w 2008 roku. Celem słownika jest rejestracja jednostek leksyki zarówno neutralnej, jak i potocznej oraz udostępnienie ich przekładów tłumaczom, studentom i użytkownikom poszukującym adekwatnych ekwiwalentów kontekstowych. Liczy on około 60 000 jednostek leksykalnych, z których znaczna część to jednostki wielowyrazowe – kolokacje, idiomy, przysłowia oraz związki potoczne – ujmowane wewnątrz haseł głównych, a nie jako odrębne artykuły hasłowe. Mikrostruktura każdego hasła zawiera informacje gramatyczne, znaczeniowe, stylistyczne oraz pełne odpowiedniki tłumaczeniowe w języku czeskim, uwzględniające odpowiedni rejestr językowy. Na uwagę zasługuje szeroki repertuar jednostek frazeologii idiomatycznej i potocznej, wśród których można przytoczyć następujące przykłady: *zrobiło się ciemno w oczach* (czes. *před očima se mi setmělo*); *rozdziierać szaty* (czes. *trhat si roucho*); *zrobić z kogoś wariata* (czes. *dělat z někoho blázna*) oraz *mieć wielkie oczy* (czes. *dělat velké oči*). Słownik zawiera również jednostki o funkcji fatycznej i emotywniej, często w postaci całych, reprodukowalnych wypowiedzi z języka mówionego. Wśród nich znajdują się zwroty takie jak *to co?* (czes. *tak co?*); *czy się jeszcze zobaczymy?* (czes. *uvidíme se ještě?*) czy *kiedy mi obiecałeś!* (czes. *vždyť jsi mi to slíbil!*), pełniące funkcje kontaktowe i ekspresywne. Pojawiają się również formy takie jak *kiedy nie jestem* (czes. *vždyť nejsem*) czy *co ty nie powiesz* (czes. *no to je mi novina*), łączące ironię z podtrzymywaniem interakcji. Pod tym względem słownik Olivy wyróżnia się na tle innych opracowań dwujęzycznych, które w znacznie mniejszym stopniu ilustrują jednostki w postaci często używanych zwrotów sytuacyjnych. Można zatem stwierdzić, że poza podstawową funkcją przekładową opracowanie to dostarcza szeregu gotowych polsko-czeskich par tłumaczeniowych w obszarze wielowyrazowych reproductów języka mówionego. Bogaty zasób leksyki ogólnej i potocznej umożliwia jego wykorzystanie w badaniach kontrastywnych oraz traktowanie go – przynajmniej częściowo – jako źródła wyodrębniania reproductów, ustalania dla nich ekwiwalentów, ich weryfikacji oraz dalszego zastosowania w opracowaniach leksykograficznych i tłumaczeniowych.

Wśród licznych prób uchwycenia czeskiej frazeologii w leksykografii przekładowej szczególne miejsce zajmuje *Wielki czesko-polski słownik frazeologiczny* pod redakcją Teresy Zofii Orłoś z 2009 roku. Stanowi jedno z najbardziej reprezentatywnych opracowań czesko-polskiej frazeologii i opiera się przede wszystkim na materiałach z opracowania *Slovník české frazeologie a idiomatiky* Františka Čermáka, wzbogaconych o współczesne źródła mówione,

medialne i internetowe. Słownik zawiera około 5000 artykułów hasłowych, z których każde opracowane jest według przejrzystego schematu mikrostruktury. Frazeologizmy lub kolokacje czeskie, często podane w kilku wariantach, opatrzone odpowiednim kwalifikatorem, polskim objaśnieniem znaczenia oraz jednym bądź kilkoma odpowiednikami tłumaczeniowymi – frazeologicznymi, kolokacyjnymi lub leksykalnymi. Tłumaczenia różnicowane są ze względu na stopień ekwiwalencji, a same hasła uzupełniono o synonimy, odwołania do innych jednostek oraz przykłady użycia zaczerpnięte m.in.: z NKJCz, literatury, prasy i internetu. Jako klasyczne opracowanie frazeologiczne, publikacja ta koncentruje się na rejestracji stabilnych jednostek językowych, takich jak idiomy, utarte zwroty, skrzydlate słowa i konstrukcje typowe dla czeszczyzny ogólnej czy potocznej. Z kolei reproduktory o funkcji fatycznej praktycznie nie występują, a wyrażenia ematywne pojawiają się sporadycznie i nie są systemowo wyodrębniane ani klasyfikowane. Do nielicznych wyjątków należy ematyw *hlava vzhůru!*, odpowiadający polskiemu *głowa do góry!*, który może pełnić funkcję ekspresywną. Podobny charakter ma zwrot *kam jsi dal(a) oči?*, tłumaczony jako *gdzie ty miałeś/miałaś oczy?*, funkcjonujący jako oceniające wyrażenie emocjonalne. Tego typu jednostki występują jednak jedynie incydentalnie, co wynika z faktu, że ich rejestracja nie należy do zasadniczych założeń słownika, choć w niektórych przypadkach możliwe jest ich wyodrębnienie. Analizowane źródło należy uznać za wartościowe opracowanie translatoryczne, zawierające bogaty zasób frazeologii i idiomatyki czeskiej wraz z odpowiednikami w języku polskim. Jego rozbudowana mikrostruktura, systematyczne podejście do ekwiwalencji oraz obecność przykładów użycia czynią go przydatnym narzędziem zarówno w praktyce tłumaczeniowej, jak i w badaniach nad frazeologią. Należy jednak podkreślić, że odsetek zarejestrowanych wielowyzrazowych jednostek o charakterze fatycznym i ematywnym pozostaje zbyt niski, by słownik mógł stanowić istotne źródło w badaniach nad tą kategorią reproduktów.

Warto wspomnieć także o jednym z najbardziej funkcjonalnych i popularnych narzędzi leksykograficznych w relacji polsko-czeskiej/czesko-polskiej, a mianowicie o *internetowym słowniku Lingea*, dostępnym bezpłatnie pod adresem <https://slovníky.lingea.pl>. Opracowanie to stanowi część szerszego projektu wielojęzycznego, realizowanego przez czeskie wydawnictwo Lingea, jednego z największych twórców elektronicznych baz leksykalnych w Europie Środkowej i rejestruje także inne pary językowe, m.in. czesko-angielską, czesko-niemiecką czy czesko-francuską. Pomimo braku precyzyjnych informacji o dacie powstania wersji czesko-polskiej, wiadomo, że projekt Lingea istnieje od wczesnych lat XXI wieku i jest stale aktualizowany. Dany słownik elektroniczny wyróżnia się przejrzystym, intuicyjnym interfejsem, przy czym wyszukiwanie odbywa się wyłącznie poprzez pojedyncze leksemy, bez

możliwości przeszukiwania fraz wielowyrazowych. Mikrostruktura każdego hasła zawiera podstawowe informacje gramatyczne, a w przypadku rzeczowników – także pełną odmianę, przy czym artykuły hasłowe zostają uzupełnione o liczne przykłady użycia, często w formie pełnych zdań, co nadaje słownikowi wyraźnie praktyczny charakter. Na szczególną uwagę zasługuje obecność licznych jednostek wielowyrazowych, takich jak frazeologizmy, kolokacje czy zwroty idiomatyczne, które występują w rozszerzonej części artykułów hasłowych. Aby jednak je odnaleźć, użytkownik musi najpierw wyszukać leksem podstawowy, ponieważ dopiero w jego obrębie pojawiają się przykłady większych całości leksykalnych, co sprawia, że identyfikacja jednostek złożonych przypomina proces ich stopniowego tropienia. Jednym z przykładów haseł, w których widoczna jest obecność jednostek fatycznych i emotywnych, jest artykuł *jak*, zawierający obok podstawowych znaczeń również następujące przykłady reproductów: *Jak se jmenujete? - Jak się pan(i) nazywa?; Jak se máš? - Co slychać?; Jak to jde? - Jak leci?; Jak chcete - Jak pan(i) sobie žyczy*. Podobne cechy można zaobserwować w hasle *myśleć*, w którym uwzględniono przykłady reproductów fatycznych bądź emotywnych, np. *Myslím, že ano/ne - Myšlę, že tak/nie; Co tím myslíte? - Co chce pan(i) przez to powiedzieć?/ Co ma pan(i) na myšli?*. Pomimo braku systemowej klasyfikacji tego rodzaju jednostek, słownik ów wyróżnia się wysoką wartością użytkową dzięki obecności licznych, aktualnych konstrukcji języka mówionego wewnątrz struktury haseł. Dzięki przejrzystej mikrostrukturze i praktycznym przykładom użycia umożliwia skuteczne tropienie jednostek wielowyrazowych, w tym także wyrażen o funkcji fatycznej i ematywnej, co czyni go narzędziem przydatnym w wyodrębnianiu i tłumaczeniu reproductów wielowyrazowych.

Na zakończenie rozważań nad aktualnym stanem leksykografii przekładowej warto zwrócić uwagę na projekt, łączący w sposób nowoczesny i spójny trzy perspektywy językowe: polską, czeską i rosyjską. Mowa o publikacji cyfrowej pt. *Lidé–Média–Komunikace. Slovník komunikačních jednotek v trojjazyčném uspořádání* z 2023 roku, opracowanej przez zespół związany z Uniwersytetem Opolskim i Uniwersytetem Ostrawskim. Słownik ten, opracowany głównie w formie elektronicznej, stanowi kontynuację i rozwinięcie idei opolskiego idiomatykonu leksykograficznego i wpisuje się w praktykę kodyfikacji oraz przekładu wielowyrazowych reproductów osadzonych w określonym kontekście komunikacyjnym. Niniejsze opracowanie rejestruje kilkaset trójjęzycznych ekwiwalentów odpowiadających sobie znaczeniowo i funkcjonalnie, zestawiając współczesne jednostki używane w komunikacji społecznej, medialnej, ekonomicznej, konsumenckiej oraz w języku nowych technologii.

Jednostki zostały uporządkowane według pięciu głównych obszarów tematycznych:

1. Dzisiejsze media analogowe i cyfrowe,
2. Komunikacja międzyludzka w dzisiejszych mediach,
3. Media jako pośrednik w interesach między ludźmi,
4. Zagrożenia i bezpieczeństwo w sieci oraz
5. Polityka komunikacyjna Unii Europejskiej (Borysowski 2024: 468).

W słowniku znaleźć można przykłady takie jak: *aukcja błyskawiczna* – *blesková aukce* – *мгновенный аукцион*; *doładaj do koszyka* – *přidat do košíku* – *добавить в корзину*; *portal społecznościowy* – *sociální síť* – *социальная сеть*; *fake news* – *falešná zpráva* – *фейк-ньюс*; *wiadomość prywatna* – *soukromá zpráva* – *личное сообщение* i wiele innych. Zgodnie z przyjętą metodą opolskiej szkoły leksykograficznej, zestawienia te mają formę zwięzłych trójjęzycznych odpowiedników przekładowych, prezentowanych bez dodatkowych komentarzy dotyczących użycia. Pomimo braku rejestracji jednostek typowych dla języka mówionego, w tym replektów fatycznych i emotywnych, jego struktura oraz metodologia stanowią cenny wzorzec dla przyszłych opracowań w tym zakresie. Trójjęzyczne zestawienie jednostek wielowyrazowych w układzie polsko-czesko-rosyjskim, przejrzystość prezentacji, forma cyfrowa oraz funkcjonalna organizacja materiału czynią z tego projektu nowatorskie dzieło na gruncie współczesnej leksykografii przekładowej. Model ten, choć skoncentrowany na języku komunikacji medialnej i specjalistycznej, powinien stać się punktem wyjścia do opracowywania analogicznych słowników w innych dziedzinach – w szczególności w obszarze jednostek fatycznych i emotywnych. Projekt ten potwierdza, że zestawianie jednostek wielowyrazowych w oparciu o kryteria tematyczne i funkcjonalne może stanowić efektywną metodę opracowywania materiału leksykograficznego perspektywie porównawczej i przekładowej.

Przedstawiona w tym rozdziale analiza wykazała, że zarówno leksykografia jednojęzyczna, jak i przekładowa, mimo długiej i bogatej tradycji, przechodzi obecnie istotne przemiany metodologiczne i funkcjonalne. Można dostrzec rosnącą świadomość konieczności rejestracji jednostek wielowyrazowych, zwłaszcza tych charakterystycznych dla języka mówionego, tj. potocznych, fatycznych czy emotywnych. Pomimo postępującej digitalizacji oraz intensywnego rozwoju lingwistyki korpusowej, zdecydowana większość współczesnych słowników (zarówno jednojęzycznych, jak i przekładowych) nadal unika systematycznego opisu tego typu jednostek, traktując je jako trudne do jednoznacznego zdefiniowania, niestabilne i nieprzystające do tradycyjnego opisu leksykograficznego. Można jednak zauważyć, że wśród współczesnych prac pojawiają się konkretne opracowania, które

przełamują dominującą tendencję marginalizacji jednostek wielowyrzowych, rejestrując różnorodne reproduktory języka mówionego oraz podejmując próbę ich kodyfikacji. Prezentowane w nich podejścia różnią się zakresem selekcji materiału, sposobem klasyfikacji i formą prezentacji, ale ich zestawienie i twórcza synteza może posłużyć jako punkt wyjścia do opracowania nowoczesnego słownika przekładowego jednostek wielowyrzowych o charakterze fatycznym i emotywnym.

ROZDZIAŁ II

Zjawisko reprodukowalności

Pojęcie reprodukowalności łączyć należy z sytuacjami wszelkiego rodzaju aktywności ludzkiej, co automatycznie przypisuje jej charakter silnie interdyscyplinarny. Nie sposób wykluczyć danego terminu i jego synonimów z obszarów nauk przyrodniczych, mówiąc o procesach powstawania nowych organizmów czy badań ekonomicznych, dotyczących odtwarzania zużytej wielkości czynników produkcji. Termin ten jest regularnie wykorzystywany w licznych obszarach badań i wiedzy m.in. w psychologii, prawie, informatyce, fotografii, malarstwie itp., w których *reprodukowalność* służy określeniu zupełnie odmiennych zjawisk, charakterystycznych dla konkretnej dziedziny nauki. Obecność tego terminu w szerokim wachlarzu dyscyplin naukowych spowodowała powstanie repertuaru nazw bliskoznacznych i pokrewnych, np. *powtarzalność*, *kopiowanie*, *imitacja*, *duplikacja*, *replikacja* czy *rekursywność*, z których każda dyscyplina czerpie według własnych potrzeb. Według Aleksandry Hirszfeld, kategoria powtórzenia jest nieodłącznym elementem struktury kultury współczesnej i należałoby mówić o niej jako pewnego rodzaju „zwrocie repetetywnym” w tej dziedzinie (Hirszfeld 2015: 8). Bez względu na szeroki zasięg pojęcia reprodukowalności i zawierania się w nim zjawisk niejednorodnych służy on przede wszystkim określeniu utraty pewnej unikatowości opisywanego tworu przy jednoczesnym zyskaniu statusu wzorca i powszechności. W taki sposób konstruuje się pewien model, niebędący tym oryginalnym, a reprezentujący pewną pochodną pierwowzoru z gotowością do dalszej replikacji i wszelkiego rodzaju przekształceń. Zachodzące w tym procesie rezultaty działań można ująć w następujące zestawienia opozycji:

‘wcześniejszy vs późniejszy’;

‘pojedynczy vs mnogi’;

‘oryginalny vs pochodny’;

‘wyjątkowy vs powszechny’;

‘macierzysty vs potomny’ (Chlebda 2020: 30).

Przywołana charakterystyka może cechować każde zjawisko bez względu na strukturę oryginału i sferę jego występowania, co umożliwia przypisanie zjawiska reprodukcji każdemu tworu materialnemu. Jednak sama reprodukowalność oprócz samej istoty duplikacji zawiera w

sobie przede wszystkim „potencjał do bycia reprodukowalnym”, a za finalizację tej cechy należy uznać jej aktualizację, czyli *akt reprodukcji*.

2.1 Reprodukowalność w języku

Do jednego z obszarów, w którym najsilniej przejawia się realizacja procesu reprodukcji odnieść należy przestrzeń komunikacji międzyludzkiej. Oczywiście jest, że każdy współczesny system językowy składa się z siatek uporządkowanych połączeń wyrazowych, które w określonym momencie niejako przeniknęły do podświadomości i w obliczu konkretnej potrzeby komunikacyjnej automatycznie odtwarzają wzorzec. Odtwarzalność w tym przypadku rozumiana jest jako umiejętność przypomnienia konkretnego połączenia wyrazowego, który w gotowej postaci (bądź wariacie) wydobyty jest z pamięci użytkownika języka, lecz nie jest tworem nowym, tworzonym od początku. Aktualizacja tego rodzaju jednostek zachodzi szczególnie w sytuacjach, kiedy ich użycie pozwala dokładnie oddać intencję wypowiedzi autora (np. wyrażenie emocji, intencji czy opinii). Potwierdza to obserwacja Julii Iwanowej, twierdzącej, iż «субъект речи находится в состоянии выбора одной из многочисленных форм, которые предоставляет в его распоряжение система языка для выражения конкретного намерения» (Иванова 2016: 155). Warto w tym miejscu przypomnieć również stwierdzenie Johna Sinclaira o tym, że język jest regulowany przez dwie równorzędne i uzupełniające się zasady: zasadę otwartego wyboru i zasadę idiomu (Sinclair 1991: 109–110). Pierwsza reguła umożliwia swobodne i jednocześnie zgodne z przyjętymi normami gramatyczno-składniowymi łączenie poszczególnych jednostek leksykalnych i konstruowanie z nich określonych wypowiedzi. Druga zaś wykorzystuje zawarte w potocznej pamięci językowej szablony, odpowiadające wymogom sytuacji komunikacyjnej. Podobną koncepcją cechują się również refleksje Piotra Müldnera-Nieckowskiego, sprowadzającego główny sens procesu wypowiedzi do rozróżnienia dwóch równoległych procedur, z których pierwszą uznać należy za odtwarzalną (dostosowaną do konkretnej sytuacji), a druga stanowi wytwór własny autora komunikatu (Müldner-Nieckowski 2007: 64).

Takie spostrzeżenie narzuca przymus obserwacji zjawiska reprodukowalności z punktu widzenia psychofizjologii, zgodnie z którą owa „gotowość” użycia środków językowych jest nierozdzielnie związana z lingwistyką kognitywną. Zgodnie z jej założeniami człowiek przechowuje w pamięci przetworzone doświadczenia językowo-sytuacyjne w postaci „śladów” użycia znaków językowych, tzw. *wzorców aktywacji*, których komponenty są ze sobą silnie powiązane. Pod pojęciem takiego śladu rozumieć należy proces i rezultat wzajemnego

oddziaływania doświadczeń werbalnych i niewerbalnych, jak i przeżyć indywidualnych czy społecznych (Швец 2005: 46). W procesie tworzenia wypowiedzi, podporządkowanej określonemu celowi, zakodowany w ludzkim umyśle algorytm aktywności mówno-poznawczej i obecność gotowych do użycia połączonych znaków językowych usuwa opozycję między świadomym a mechanicznym użyciem połączeń. Zgodnie z teorią Michaiła Dymarskiego, «в процессе порождения высказывания некоторый смысл, который подлежит выражению, воплощается в результате оживления существующих в мозгу человека ассоциативных связей, неоднородных комбинаторных моделей, узловые точки которых представляют слова и сочетания слов, избираемые в соответствии с конкретной областью обозначения» (Дымарский 2013: 325–326). Metaforyczne ujęcie przechowywania połączeń językowych przedstawia również Emanuel Szubin, lokalizując w pamięci ludzkiej dziesiątki tysięcy matryc znaków różnych rzędów, cechujące się strukturą uporządkowaną i są połączone poprzez relacje formalne, funkcjonalne, paradygmatyczne i syntagmatyczne. Za najsilniejsze uważa ostatnie, kumulujące wiedzę o najczęstszych typach związków, w jakich dany znak był używany (Szubin 1974: 92–93). Kognitywną podstawą odtwarzalności może być również obraz, rozumiany jako wynik maksymalnej kondensacji pierwotnego wyobrażenia z aktualizacją obrazu mentalnego, tworzącego się na podstawie emocjonalno-wizualnego przetworzenia informacji przez jeden lub więcej kanałów sensorycznych (Красных 2016: 126). Z kolei sama umiejętność przetrzymywania w pamięci ogromnych połączeń językowych uwarunkowana jest sposobnością mózgu reagować analogicznie na znane świadomości wpływy środowiska i traktować je jako jednolity system, a nie jako nowe oddzielne sytuacje (Абакимова 2020: 15).

Dla przedstawionych wyżej założeń fundamentalne znaczenie ma zjawisko pamięci zbiorowej, którą można definiować jako obrazy przeszłości aktywowane w procesie komunikacji społecznej w celu zaspokojenia kulturowych i politycznych potrzeb teraźniejszości (Czachur 2018: 11). Wspomniane obrazy przeszłości, wśród których wyróżnić można różnorodną liczbę kodów (werbalnych i niewerbalnych) przekazywane są poprzez praktyki pamięci, a sama pamięć zbiorowa istnieje bezpośrednio dzięki komunikacji ustnej i pisemnej. Zaprezentowany zespół wyobrażeń o przeszłości jest niewątpliwie bytem ujęzykowanym, przy czym język zgodnie z definicją Wojciecha Chlebdy, pełni tu dwie równorzędne funkcje: Formująca siła werbalności [...] jest dwuwektorowa i działa w dwu przeciwnych kierunkach: „do wewnątrz” kiedy kształtuje treść (zawartość) pamięci samej w sobie (wektor dośrodkowy) oraz „na zewnątrz”, kiedy kształtując oddziałuje na postać późniejszych relacji tej treści (wektor odśrodkowy) (Chlebda 2018: 60).

Ujęzykowiony obraz przeszłości wyrażany zostaje poprzez gotowe schematy komunikacyjne, a gatunek mowy z kolei wpływa na kształtowanie samej interpretacji przeszłości (Markowski 2006: 159). W kręgu przedstawionych badań zasadnie więc mówić o tzw. *pamięci komunikatywnej*, do której odnieść należy cztery następujące po sobie fazy: zapamiętywanie, pamiętanie, przypominanie i zapominanie. To właśnie fazy pamiętania i przypominania dotyczy cały proces, związany z aktem reprodukcji, w przypadku którego brak regularnego użycia jednostki językowej skutkuje stopniowym przechodzeniem do fazy zapominania. Reprodukcję należy uznać za warunek konieczny do realizacji skutecznego transferu treści społeczno-kulturowych, w których uczestniczą wytwory pamięci (magazynujące i funkcjonalne), stanowiące jej praktykę. Istotnie pamięć zbiorowa jest odpowiedzialna za przekazywanie pewnych zachowań językowych czy sposobu użycia trwałych połączeń wyrazowych stosowanych w określonych sytuacjach komunikacyjnych.

W każdym akcie mowy człowiek zajmuje bądź postawę analityczno-ekspresywną, bądź postawę naśladowczo-przystosowawczą (Lewicki 1976: 22). Tym samym każdy komunikat stanowi pewnego rodzaju stadium pośrednie między odtwarzaniem tego, co zostało wcześniej zapożyczone a autorskim tworzeniem wypowiedzi. Olga Dołgowa w swoich badaniach zauważyła, iż nie każda wypowiedź wiąże się z powstaniem nowego tworu oryginalnego. Każdorazowy akt mowy przyczynia się do zwiększenia ogólnej masy wypowiedzeniowej, aczkolwiek ich zawartość informacyjna może być tożsama i nie tworzy wtedy bytu unikatowego (ДоЛГОВА 1980: 165). Zjawiska te nie tylko wzajemnie się uzupełniają, ale wykluczają swój pierwotny charakter dychotomiczny, tworząc trzelementowy ciąg zjawisk: *odtworzenie czyjejś mowy – stworzenie tworu nieoryginalnego – stworzenie tworu oryginalnego*. Z kolei Uriel Weinreich przedstawia szkicowy model procesu generowania zdania powstającego w wyniku zarówno kompilacji prostych elementów leksykalnych przy zachowaniu obowiązujących reguł gramatycznych, jak i wyboru gotowych struktur złożonych (formuł, powiedzeń, porzekadeł itp.). Obie te drogi tworzą harmonię i konstruują każdorazowe wypowiedzenie (Weinreich 1969: 23–81). Za synonim struktur złożonych uznać można grupy wyrazowe, które wbrew dotychczasowej opinii językoznawców nie są każdorazowo generowane przez osobę mówiącą, lecz reprodukowane w postaci gotowych segmentów tekstowych (Wójcik 1983: 15). Niezwykle trudnym zadaniem jest ustalenie, w jakim stopniu człowiek mówiący wykorzystuje w mowie trwałe i odtwarzalne połączenia wyrazowe, a w jakiej części są to jednostki autorskie i indywidualne, tworzone ze „swobodnych” elementów języka (np. pojedynczych wyrazów).

Według Andrzeja Marii Lewickiego, twórcy założeń frazeologii syntaktycznej, połączenia utrwalone odwołują się do pamięci odbiorcy i mogą być przez niego traktowane całości. Badacz wyróżnia połączenia regularne, umożliwiające słuchaczowi odtworzenie procesu generowania połączenia z pamięci, bądź nieregularne, uniemożliwiające zrozumienie komunikatu z powodu braku skojarzeń z poprzednimi doświadczeniami komunikacyjno-sytuacyjnymi (Lewicki 2003: 38–39). Nawet wspomniane przez Lewickiego całościowe ujęcia połączeń wyrazowych mogą być przez użytkowników odbierane jako zdolne do dalszego podziału i zyskujące tym samym zdolność do łączenia się z innymi jednostkami, zachowując najczęściej swoje znaczenie. Elena Kubrjakowa odnotowuje, iż w pewnym sensie można mówić o priorytecie całościowego postrzegania, przechowywania i wydobywania z pamięci gotowych jednostek nominacji, nawet tych, które dla samych mówców łatwo rozkładają się na elementy i segmentują (Кубрякова 1991: 136). Potwierdzają to również obserwacje Wojciecha Chlebdy, który związek frazeologiczny traktuje jako twór formalnie podzielny leksykalnie i zapamiętywany w dwóch postaciach: jako całość leksykalna z założonym znaczeniem i pod każdym z jej elementów traktowanym oddzielnie (Chlebda 2020: 39). Podobne stanowisko prezentuje również Danuta Buttler, która przypisuje frazeologizmowi nie tylko znaczenie realne, ale również strukturalne, stanowiące sumę wartości somatycznych jego komponentów (Buttler 1982: 52). W takim ujęciu jednostki frazeologiczne zapamiętywane zostają nie tylko jako całości nominacyjne, ale i komponenty werbalnie rozkładalne, cechujące się niezależnym potencjałem asocjacyjnym.

Aktualizacja reprodukowalności w ludzkiej komunikacji jest zarazem sposobem organizowania całego toku wypowiedzi, a repetytywność określonych ciągów wyrazowych zmierza do powtarzania tego, co zostało stworzone przed samym aktem mówienia. Potwierdzają to XIX-wieczne obserwacje rosyjskiego literaturoznawcy, Michaiła Bachtina, który zakładał, że „każda wypowiedz to ogniwo niezwykle złożonego łańcucha innych wypowiedzi. [...] Wypowiedzi nie zachowują się obojętnie wobec siebie ani do siebie nie przystają: są sobie świadome i nawzajem odzwierciedlają się w sobie. [...] Każda wypowiedź pełna jest ech i pogłosów innych wypowiedzi, z którymi uczestniczy we wspólnocie obcowania językowego w danej sferze” (Bachtin 1986: 361, 390). Całą komunikację międzyludzką stanowią fundamenty mechanizmów formułiczności, a zapamiętywanie i powtarzalność określonych wiązek sensów służy zapewnieniu ciągłości kulturowej i kondensacji zdobytej wiedzy. Zjawisku reprodukowalności językoznawcy na przełomie wieków poświęcili немало uwagi, aczkolwiek obszar tego, co rzeczywiście zawiera to zjawisko i jakie twory językowe można uznać za odtwarzalne nadal pozostaje kwestią dyskusyjną.

Reprodukcja według Lewickiego to „świadome powtórzenie połączenia wyrazów, które było już użyte w innej sytuacji komunikatywnej. Świadome powtórzenie to takie, przy którym reprodukowany fragment jest strukturalnie, znaczeniowo lub stylistycznie niespójny z resztą wypowiedzi. Niespójność ta jest sygnałem wyodrębnienia go przez nadawcę i powinna zmusić odbiorcę do zwrócenia uwagi na dany fragment wypowiedzi” (Lewicki 2003: 120–121). W tym miejscu należałoby przedstawić pierwszą podjętą przez badacza próbę typologii reprodukcji, do których odnieść można reprodukcję indeksowaną i nieindeksowaną. Pod pierwszym pojęciem rozumiane jest wierne odtworzenie fragmentu bądź całości czyjejś wypowiedzi, najczęściej w postaci cytatu. Nierzadko można mówić o reprodukcji z innowacją, w której mówiący mniej lub bardziej świadomie dokonuje zmian w strukturze ciągu wyrazowego. Niewątpliwie zatem można utożsamić reprodukcję indeksowaną ze sferą użycia skrzydlatych słów, wśród których często dochodzi do wymiany elementów z powodu błędów użytkownika języka. Z kolei przy reprodukcji nieindeksowanej nie mamy do czynienia z bezpośrednim odwołaniem do konkretnego autora czy sytuacji – te wielowyrazowe jednostki są „dobrem powszechnym ogółu” i odnoszą się do pamięci zbiorowej danej wspólnoty językowej (Chlebda 2010b: 16). Zilustrowaniem przykładów takich jednostek mogą być wyrażenia typu *Co to, to nie; W żadnym wypadku!; Proszę za mną*. Przedstawioną klasyfikację Wojciech Chlebda dostosowuje do założeń nowoczesnej koncepcji frazeologii pragmatycznej, zwanej frazematyką, której poświęcona zostanie dalsza część pracy. Badacz wyodrębnia replikacje nieintencjonalne (mechaniczne i nieświadome powtórzenia ciągów werbalnych) i replikacje intencjonalne, które z kolei dzieli na trzy grupy: odtwórcze, regularne i twórcze. Obszar pierwszego zbioru jest ogromny, a odnieść do niego można napisy tablic informacyjnych, komunikatów dworcowych czy napisów ostrzegawczych. Ich użycie przeważnie wynika z potrzeby ekonomii języka, przy czym same wypowiedzi są w pełni zamierzone, tworzone według określonych klisz językowych i składające się z niewielkiej liczby wyrazów. Replikacje regularne Chlebda stawia na równi z definicją standardowego związku frazeologicznego: „ciąg macierzysty pozostaje składową systemu danego języka, ciąg pochodny jest jego przywołaniem” (Chlebda 2020: 36). Reprodukacja w tym przypadku może być kształtowo i semantycznie zbieżna z oryginałem, bądź stanowić jeden z jego wariantów i charakteryzować się niedokładnością. Intencjonalność może być w takim przypadku zasygnalizowana użyciem formuł wskaźnikowych, jak np. *jak dawniej mawiano*. Grupę ostatnią stanowią indywidualno-autorskie transformacje, do których badacz odnosi nie tylko język poetów i pisarzy (jak przyjęto w tradycyjnej frazeologii), ale również zjawiska przekształceń utartych jednostek wyrazowych przez zwykłych użytkowników języka. Takie spojrzenie na naturę reprodukowalności narzuca

wniosek, iż każdy ciąg wyrazowy może być dla określonych celów zreprodukowany, a każda reprodukcja może być poddana dalszej transformacji. Co więcej, wspomniany fenomen odznacza się odnawialnością określonego wyrażenia zarówno bezpośrednio jak i w sposób odwleczony w czasie i przestrzeni, o czym przekonuje Wiktoria Krasnych (Красных 2016: 122). Różne ujęcia terminu reprodukowalności narzucają dogłębną analizę klasycznego rozumienia tego pojęcia w lingwistyce, jak i scharakteryzowanie źródeł i zmian w badaniach nad tym zjawiskiem.

2.2 Reprodukacja we frazeologii

Tradycyjne ujęcie reprodukowalności wiązało się z przyjętą dychotomią pojęć język – mowa, przypisując tym samym wymienioną cechę tylko określonym jednostkom języka, jak np. idiomom czy frazeologizmom. Odtwarzalność na równi z metaforycznością, ekspresyjnością czy obrazowością są cechami konstytutywnymi frazeologizmów, ich cechą kategorialną i jedną z przyjętych metod wyodrębniania utrwalonych połączeń wyrazowych. Takie spojrzenie naturalnie wymusiło eliminację innych jednostek, które wszystkich kryteriów narzuconych frazeologizmom nie spełniają, co spowodowało ich permanentne pomijanie w opracowaniach leksykografii jedno- czy dwujęzycznej. Próżno szukać we współczesnych słownikach wielowyrzowych odtwarzalnych formuł, jak np. *Żeby mi to było ostatni raz!*; *A Pan do kogo?*; *Co Pani nie powie!*, które przez tradycyjne paradygmaty frazeologiczne były marginalizowane i pozostawały poza polem badawczym językoznawców. Wielokrotnie pomijany fakt interdyscyplinarności frazeologii, która stanowi szerokie pole badawcze dla zagadnień lingwistycznych, społeczno-kulturowych czy filozoficznych doprowadził do zamknięcia zjawiska reprodukowalności wyłącznie w jej granicach.

Przed przejściem do dalszych rozważań warto przytoczyć cytaty Andrzeja Marii Lewickiego, który niejako wprowadzi czytelnika do omawianej problematyki i będzie stanowił myśl przewodnią w analizie sformułowanych dotychczas teorii frazeologicznych - „Nie ma jednej frazeologii, lub jeśli kto woli, frazeologia jest nauką wieloaspektową. Każda z odmian frazeologii zakłada inny sposób patrzenia na język i inne stawia sobie cele. Tylko wyraźne uświadamianie sobie punktu widzenia i celów może przynieść tej nauce trwałe osiągnięcia” (Lewicki 1974: 151). Jednym z nurtów, który zapoczątkował rozwój kanonicznej teorii frazeologicznej były koncepcje Wiktora Winogradowa, który stworzył potoczne (i do dziś istniejące) wyobrażenie formy jednostki frazeologicznej, zyskując uznanie środowiska naukowego i licznych kontynuatorów. Była to pierwsza próba usystematyzowania i

szczegółowej charakterystyki zbiorów frazeologicznych języka rosyjskiego podzielonych na podstawie kryterium semantycznego na trzy grupy:

- фразелогические сращения,
- фразеологические единства,
- фразеологические сочетания. (Виноградов 1977: 121)

W klasyfikacji na plan pierwszy wysuwa się stopień spójności znaczeniowej frazeologizmu, tj. aspekt korelacji między globalnym znaczeniem całej jednostki a znaczeniem jej elementów. Przy przyporządkowaniu konkretnej jednostki do określonej grupy posługiwano się głównie metodami intuicyjno-opisowymi, co poskutkowało kategoryzacją jednostek „semantycznie przesuniętych” na związki stałe, inne zaś uznano za związki luźne. W tym miejscu należy zaznaczyć, że nawet współczesna lingwistyka nie operuje wypracowanymi i skutecznymi metodami określenia stopnia semantycznego zespolenia komponentów. Niemniej jednak przedstawiony zarys koncepcji Winogradowa stał się wzorcową teorią frazeologii ubiegłego wieku, wyznaczająca wyraźną granicę między stałymi związkami wyrazowymi (frazeologizmami) a związkami luźnymi, pozbawiając te ostatnie statusu frazeologicznego.

Nieco inną koncepcję, przygotowującą podwaliny pod współczesne rozumienie frazeologii, przedstawił Jewgienij Poliwanow, który w krótkim ujęciu skonstruował podstawowe wyznaczniki nowego ujęcia frazeologii. Za punkt wyjścia dla swoich teorii przyjął umieszczenie jednostki frazeologicznej w konkretnej sytuacji – sytuacji wypowiedzi, a nie jak dotychczas uważano w izolacji. Potwierdzają to jego obserwacje, zawarte w jednym z jego artykułów: „W zwykłych warunkach mówienia trzeba wykluczyć z aktu myślenia językowego jako takiego tworzenie symbolu, który miałby wyrażać pewne wyobrażenie pozajęzykowe: Symbol ten jest już gotowym (dla przykładu, mówiący nie musi w odpowiedzi na pytanie *„Czy pił pan herbatę?”* tworzyć nowej frazy *„Tak, dziękuję”* – Fraza ta jest już gotowa w asortymencie jego myślenia słownego, frazeologicznego jako określony, stały symbol dla wyrażenia danej wiązki idei” [tłum.- Wojciech Chlebda] (Поливанов 1968: 59). Jego spojrzenie na naturę jednostki frazeologicznej uwzględniało zarówno aspekt psycholingwistyczny i pragmatyczny, jak i po raz pierwszy nadrzędną rolę nadawcy wypowiedzi – akt mowy rozumiany jako proces przejścia od intencji do form ich wyrażania w języku.

W taki sposób uformowały się dwa przeciwstawne nurty, gdzie kierunek poliwanowski wychodzi od tekstu do postaci słownikowej, a zwolennicy Winogradowa skupiają się na systemowych postaciach jednostek frazeologicznych, pozbawionych umiejscowienia w kontekście sytuacji. Wykorzystanie wypowiedzi w tradycyjnym ujęciu frazeologicznym służy ilustracji przykładu użycia, a nie samej analizy jednostki, zanurzonej w otoczeniu tekstowym.

Niejednorodność ogromnego zbioru jednostek i brak wyodrębniania jednej wspólnej dla nich cechy dodatkowo komplikował fakt wprowadzenia przez Nikolaja Szanskiego czwartej grupy frazeologizmów, tzw. wyrażeń frazeologicznych, uwzględniającej jednostki o stałym składzie leksykalnym i całkowicie motywowanym znaczeniu globalnym (Chlebda 2003: 22). Chaos różnych podejść metodologicznych, założeń i sprzecznych teorii zaowocował wyodrębnieniem dużej liczby kryteriów, określających status jednostki frazeologicznej. Jednak najczęściej wspominanymi cechami frazeologizmów uczyniono odtwarzalność, stałość składu leksykalnego, globalność nominacyjną, obrazowość, idiomatyczność, ekspresyjność, metaforyczność i wreszcie rozpowszechnienie społeczne. Takie ujęcie narzuciło odgraniczenie zbioru frazeologizmów od całego repertuaru porzekadeł, przysłów, idiomów itp., traktując je tym samym jako związki luźne. Dopiero myśl Poliwanowa zdaje się łączyć wszystkie byty językowe i sprowadzać je do najważniejszego wyróżnika: odtwarzalności w określonej sytuacji, przydając jednostkom wartość pragmatyczną. Zaprezentowaną ideę wykorzystywali w swoich pracach Szanskij czy S.G. Gawrin, jednak dopiero Georgij Piermiakow zrównał frazeologizmy z całą klasą jednostek o podobnych cechach, nominując je ogólną nazwą klisz językowych, czyli gotowych wyrażeń odtwarzalnych.

W tradycyjnej frazeologii polskiej, którą łączyć należy przede wszystkim z nazwiskiem Stanisława Skorupki, mamy do czynienia z pewnego rodzaju stadium pośrednim pomiędzy koncepcjami dwóch rosyjskich lingwistów. Kryterium podziału frazeologizmów pozostawało kryterium semantyczne, jednak wyodrębnione przez Skorupkę związki stałe, łączliwe i luźne nie w pełni odpowiadały podziałowi Winogradowa. Między związkami luźnymi a фразеологическими сочетаниями nie było zgodności co do ich definicji – w сочетаниях znaczenie jednego ze składników jest znaczeniem związanym, a w ujęciu Skorupki związki luźne należałoby traktować jako związki tworzone każdorazowo doraźnie o niescalonym znaczeniu. Tym samym Skorupka uznaje wyraz za fundamentalny element frazeologizmu, wprowadza połączenia wyrazowe w orbitę frazeologii, a jednocześnie niekonsekwentnie wskazuje ich miejsce w zakres składni i semantyki. Bez względu na nieścisłość wymieniona teoria, spowodowała ona przełom w dotychczasowym rozumieniu jednostki frazeologicznej, o czym świadczą przede wszystkim koncepcje Andrzeja Marii Lewickiego i Andrzeja Bogusławskiego.

Pierwszy z nich, opracowując zasady frazeologii syntaktycznej, nazwał swoją ideę frazeologią odbiorcy, która poświęcona zostaje analizie stereotypów i związków wyrazowych, reprodukowanych z pamięci. Z kolei główne postulaty Andrzeja Bogusławskiego zawierają się w jednej z jego najważniejszych koncepcji, wyrażonej cytatami: „poszukiwanie i

wyodrębnianie wielowyrazowych jednostek języka jest centralną kwestią lingwistyczną” i „zadaniem głównym językoznawstwa” (Bogusławski 1989: 18–19). Jednostki wielowyrazowe, zdaniem badacza, „idą w grube miliony, w przeciwieństwie do jednostek jednowyrazowych, których jest kilkadziesiąt tysięcy, a nie są odnotowywane przez źródła podstawowe”. Poprzez odrzucenie opozycji frazeologia – leksyka wprowadzone zostały przez badacza następujące pojęcia: „jednostka języka” (czyli gotowe i reprodukowalne całości leksykalne) i „produkty językowe” (kompozycje tworzone drogą regularnego łączenia jednostek). Swoją koncepcją zakwestionował dotychczasowe rozumienie zakresu jednostek, które powinny być uwzględniane w słownikach frazeologicznych, a wyznacznikiem „jednostki języka” stała się jej „gotowość” i odtwarzalność w mowie, bez względu na jej strukturę. Połączenia wyrazowe, typu *no już dobrze; ach, tak; poważnie mówisz?; gdzie tam!* zyskały niespodziewanie status jednostek frazeologicznych, co spotkało się ze stanowczym sprzeciwem ze strony zwolenników tradycyjnego definiowania frazeologizmu.

Nowatorskim spojrzeniem na naturę frazeologizmów na gruncie literatury czeskiej odznaczają się prace Františka Čermáka, który odchodzi od tradycyjnego ujęcia jednostki frazeologicznej, stwarzając tym samym miejsce dla nowej dyscypliny – *frazeologie a idiomatika* (Čermák 2007). Dodanego obszaru autor zalicza wszystkie jednostki wielowyrazowe obdarzone znaczeniem, charakteryzujące się ustaloną kombinacją form wyrazowych i tzw. *anomalią*, czyli zjawiskiem unikalnego i nielogicznego łączenia poszczególnych jednostek, które w określonym zestawieniu nabierają konkretnego znaczenia. Decyzja o sformułowaniu nowej dziedziny uzasadniona zostaje zbyt wąskim obszarem jednostek uwzględnianych przez tradycyjną frazeologię, która wymusza szereg kryteriów konstytutywnych dla ich wyodrębniania, takich jak np. obrazowość, metaforyczność czy niepodzielność. Ponadto w czeskiej literaturze termin *frazeologie* jest szeroko stosowany w zupełnie innym znaczeniu, w szczególności dla opisanie nieokreślonych zbiorów praktycznych formuł i zwrotów, używanych głównie w środowiskach zawodowych. Odtwarzalne jednostki wielowyrazowe, takie jak porzekadła, przysłowia, zwroty etykietalne czy aforyzmy były rejestrowane w sposób rozproszony w rozmaitych opracowaniach leksykograficznych bądź tworzono dla nich odrębne słowniki jednojęzyczne. Przyjęcie terminu *Frazeologie a idiomatika* otwiera szeroki leksykon idiomów językowych, wśród których znajdują się wyrażenia, typu *otevřít dokořán; být hin; Že by?; To zrovna!; co kdyby; v souvislosti s něčím* itp. Badanie idiomów odbywa się na wszystkich poziomach (syntagmatycznych i paradygmatycznych) i wykracza poza ramy samego językoznawstwa, odwołując się do wiedzy z zakresu innych nauk przyrodniczych czy humanistycznych. Wyszczególnienie funkcji denominacyjnej,

strukturalnej czy pragmatycznej idiomów narzuciło ich obserwację i analizę wyłącznie w określonym kontekście tekstowym, zawierającym elementy wartościujące i symboliczne.

System językowy w ujęciu Čermáka dostarcza użytkownikowi gotowych predykatów wielowyrazowych, służących nie tylko do nominacji, lecz także stanowiących uzupełnienie języka naturalnego. Jednocześnie charakteryzują się one zdolnością do dalszego podziału, tworząc tzw. *pólidiony* – ‘jednostki zdolne do znacznej łączliwości elementów w określonym kontekście’. Zwrócenie uwagi na funkcję tekstową jednostek pozwoliło dostrzec silną wariantywność kombinacji wielowyrazowych, których najbardziej reprezentatywne analizy odbywają się na materiale języka mówionego, a nie jak dotychczas pisanego. Po raz pierwszy w krąg rozważań frazeologicznych wprowadzone zostały całe jednostki zdaniowe, jak np. *Den se dlouží; Koukej mazat!; Tak se měj!*, których regularna powtarzalność w tekstach i względna stabilność komponentów pozwoliła na ich kodyfikację słownikową w ramach nowej dyscypliny. Rezultatem wieloletnich prac Františka Čermáka stał się pięciotomowy *Slovník české frazeologie a idiomatiky*, składający się z następujących części: 1. *Přirovnání*, 2. *Výrazy neslovesné*, 3. *Výrazy slovesné*, 4. *Výrazy větné* oraz 5. *Onomaziologický slovník*. Nowoczesne ujęcie pomijanych dotychczas jednostek pozwoliło zarejestrować ogólnonarodowy zasób idiomatyki i frazeologii języka czeskiego, których analiza odbywa się na materiale korpusowym wraz z przytoczeniem ekwiwalentów obcojęzycznych dla czterech języków (angielski, niemiecki, francuski i rosyjski). Szczególne *novum* w językoznawstwie prezentuje ostatnia część słownika *Výrazy větné*, rejestrująca przykłady pełnych formuł zdaniowych stosowanych w określonych sytuacjach komunikacyjnych. Skodyfikowane jednostki są rozpatrywane jako stabilne jednostki frazeologiczne, które w tradycyjnym ujęciu były traktowane jako wypowiedzi tworzone każdorazowo z poszczególnych elementów systemu językowego.

Powszechnie znaną definicję frazeologizmu można ująć w następującym zdaniu: „frazeologizm jest to utarte połączenie wyrazów, odtwarzane w mowie w gotowej postaci, o znaczeniu metaforycznym nie wynikającym z sumy znaczeń składników, obrazowe i wzmacniające ekspresyjność wypowiedzi” (Chlebda 2003: 31). Silny wpływ nurtu Winogradowa ukształtował stereotypowe pojęcie frazeologizmu, wprowadzając na grunt badań lingwistycznych kanon jednostek wzorcowych i odrzucając jednocześnie inne połączenia i nadając im status swobodnego połączenia wyrazów. Wyraźnie wprowadzona dychotomia odrzucała wcześniej inne kompozycje odtwarzalne, które stanowią ich odmiany, warianty bądź swobodne połączenia słów, np. *традиционные сочетания* (A.I. Smirnickij), *стильные речения* (F.F. Fortunatow, A.A. Szachmatow), *составные термины* (W. W. Winogradow),

фразеологические выражения (N.M. Szanskij), *лексиколитованные сочетания* (S.I. Ożegow), *фраземы или фразеоматические сочетания* (N.A. Amosowa, I.P. Susow, A.W. Kunin), *синлексы* (G. I. Klimowskaja), *номинемы* (W.M. Nikitjewicz), *коллокации* (E.G. Borisowa) (Абакумова 2020: 13–14). We współczesnym ujęciu danej dyscypliny podejścia systemowe i tekstowe wzajemnie się uzupełniają, odwołują się m.in.: do wiedzy teorii aktów i gatunków mowy, semantyki komunikacyjnej, czyli do osiągnięć pragmatyki językoznawczej. Źródłem i polem prowadzenia badań stał się tekst i indywidualna płaszczyzna wypowiedzi, w której najważniejszymi aspektami stały się analizy osoby mówiącej, warunków komunikacji, pojęcia odtwarzalności i zależności między mówieniem a działaniem. Jak już zostało wspomniane, Andrzej Maria Lewicki poprzez rozpracowanie zagadnień teorii syntaktycznej wprowadził pojęcie frazeologii odbiorcy, akcentując tym samym konieczność zwrócenia uwagi na nadawcę komunikatu, który z czasem stał się głównym centrum zainteresowania nowego paradygmatu frazeologii teoretycznej, czyli **frazematyki**.

2.3 Ogólne założenia frazematyki

Wszelkie koncepcje frazematyczne należy łączyć z Wojciechem Chlebdą, który w swojej monografii *Elementy frazematyki* z 1991 roku przedstawia obszerny zarys działania człowieka jako istoty społecznej w procesie komunikacji. W zaprezentowanej przez badacza koncepcji człowiek mówiący w określonym celu i w konkretnej sytuacji komunikacyjnej stanowi punkt wyjścia dla dalszych rozważań. Zagadnieniem głównym nowego paradygmatu staje się akt międzyludzkiej komunikacji, który złożony jest zarówno z wypowiedzi twórczych, jak i odtwórczych, zwanych jednostkami reprodukowanymi, czyli *frazemami*. Można powiedzieć, że frazematyce przyświeca idea, skonstruowana przez Jurija Karaulowa: „za każdym tekstem stoi system języka i osoba mówiąca” (Карaulов 1989: 3–9). Pojęcie odtwarzalności nabiera tu nowego sensu – jest ona rozpatrywana w odniesieniu do konkretnej sytuacji, nadawcy i jego intencji, a powtarzalność ciągu wyrazowego służy zwerbalizowaniu określonego potencjału treściowego w analogicznych warunkach sytuacyjnych. Frazem traktowany jest nie tylko jako nośnik treści semantycznej, ale także jako bodziec do wywołania konotacji, będących nieodłącznymi elementami zawartości tekstu.

Dlatego też przedstawiona analiza nie jest poświęcona wypowiedziom utworzonym z doraźnych elementów, a jednostkom z pamięci odtworzonym, cechującym się strukturą wielowyrazową i dowolnym gatunkiem mowy. Nowy nurt reprezentuje sobą pierwszą próbę przypisania różnorodnym jednostkom cechy podstawowej – odtwarzalności, a cechy typowe

dla związków frazeologicznych (ekspresyjność, metaforyczność, obrazowość itp.) traktowane są jako własności fakultatywne, a nie obligatoryjne dla ich wyodrębniania. W przeciwieństwie do frazeologii pragmatycznej, która rozszerza je do setek tysięcy bądź nawet milionów, dotychczasowy paradygmat frazeologiczny ze względu na rygorystyczne kryteria kwalifikacji zawężał pole obserwacji do kilkunastu tysięcy jednostek. Frazematyczna zasada „od intencji i treści ku formom ich wyrażania” wyklucza uwzględnianie w analizach wszelkich ograniczeń formalno-strukturalnych. Nadrzędnym terminem swojej koncepcji Chlebda uczynił określenie, które choć występowało już w zagadnieniach *stricte* frazeologicznych, zyskuje teraz zupełnie nowe znaczenie: **frazem** jest to znak językowy, stanowiący nazwę potencjału treściowego, którą w danej sytuacji dla wyrażenia tego potencjału mówiący przywołuje (odtwarza) w charakterze jako względnie stałego symbolu. Zbiór takich znaków danego języka nazwał **frazematyką** (Chlebda 2003: 52).

Konieczność stworzenia nowego pojęcia wynikała ze zbyt trwałych asocjacji, które zakorzeniły się w ludzkiej świadomości w związku z tradycyjnym postrzeganiem frazeologii i prototypowego związku frazeologicznego. Uczynienie z odtwarzalności jednego kryterium konstytutywnego frazematyki umożliwiło włączenie do obszaru badań ogromnej liczby faktów językowych, które dotychczas nie były rozpatrywane pod kątem możliwości nadania im statusu frazeologicznego. Koncepcja frazematyczna dąży do zniwelowania różnic i podziałów między jednostkami leksykalnymi, poszukując dla nich ujednoliconych metod analizy. Do repertuaru wspomnianych frazemów należeć więc będą m.in:

Gatunek	Przykład
1. Frazeologizmy klasyczne	<i>puszka Pandory; kozioł ofiarny; biały jak ściana; stajnia Augiasza</i>
2. Przysłowia	<i>cel uświęca środki; co kraj, to obyczaj; baba z wozu, koniom lżej</i>
3. Formuły gatunkowe:	<i>Nawiązując do poprzedniej rozmowy...; z wyrazami szacunku</i>
4. Formuły etykietalne:	<i>Panie i Panowie! Przepraszam uprzejmie, ale...; Wszystkiego najlepszego!</i>
5. Elementarne formy folkloru:	<i>Czary mary; hokus pokus</i>
6. Komunikaty:	<i>Prosimy zachować ostrożność; Koniec trasy</i>

7. Skrzydlate słowa	<i>Jestem za, a nawet przeciw; Nie chcę, ale muszę</i>
8. Hasła i slogany	<i>Teraz albo nigdy! Prawie robi wielką różnicę; Takie rzeczy tylko w Erze.</i>
9. Teksty użytkowe	<i>Świeżo malowane; Przejdźcie drugą stroną ulicy, Zakaz wjazdu</i>
10. Tytuły:	<i>„Rozmowy nocą”; „07 zgłoś się”; „Lepiej późno, niż później”</i>
11. Terminy złożone i terminologizmy:	<i>Świńska grypa; komputer stacjonarny; wojska obrony przeciwlotniczej</i>
12. Operatory metajęzykowe i metatekstowe	<i>Jak to się kiedyś mówiło...; Jak mawiają na Śląsku...; Jak mówią sportowcy...;</i>
13. Wyrażenia funkcyjne	<i>A jeszcze do tego...; Żeby tego było mało...; a na dodatek...</i>
14. Wyrażenia pragmatyczne	<i>A co ci do tego?; A nie mówiłem!; No co ty nie powiesz?; A to ci heca!</i>

Można zauważyć, że paradygmat frazeologii pojmuję zasoby języka jako zbiór jednostek rozmaitych, przybierających zarówno formę jedno-, jak i wielowyrazową. W centrum zainteresowania frazematyki pozostają jednak te drugie, gdyż jednostki jednowyrazowe zostały już licznie odnotowane przez leksykografię jedno- i dwujęzyczną, w odróżnieniu od odtwarzalnych grup wyrazowych, których stopień kodyfikacji słownikowej można ocenić jako skrajnie niski. Przedstawiona tabela nie rejestruje całej różnorodności rozpatrywanych frazemów, a służy zilustrowaniu przykładów jednostek od tej pory zyskujących status frazeologicznych.

2.4 Sytuacyjność we frazematyce

Zjawisko sytuacji stanowi czynnik obligatoryjny dla zaistnienia nadawcy i odbiorcy, konkretnej wypowiedzi i wygenerowania określonego celu komunikacji. Nawet w warunkach zwracania się do samego siebie, człowiek bardziej lub mniej świadomie wytwarza sytuację komunikacyjną, w której sam jest zarówno adresatem jak i odbiorcą informacji. Zagadnienia frazematyczne można usytuować w triadzie język – myślenie – działanie, co przydaje nowemu paradygmatowi frazeologii charakter zarówno językowy, jak i ekstralingwalny, związany ze światem psychiki ludzkiej. W tym ostatnim mamy do czynienia z zachowaniem człowieka w

sytuacjach stabilnych (odtwarzalnych/powtarzalnych), jak i sytuacjach zupełnie nowych. Według Tadeusza Tomaszewskiego, sytuacje stabilne narzucają człowiekowi zachowanie „reaktywne”, tj. bierne poddawanie się napływowi informacji, które wzmacnia odłożone w pamięci ślady doświadczeń wcześniejszych. Tworzą one stabilny obraz rzeczywistości, wytwarzając konsonans poznawczy, w przeciwieństwie do informacji nowych, przyczyniających się do destabilizacji znanego obrazu i wywoływaniu dysonansu poznawczego (Tomaszewski 1984: 5–35).

Frazematyka rozpatruje generowanie wypowiedzi jako procesu wytwarzania w mózgu określonego potencjału treściowego, którego elementy łączą się najpierw poprzez zautomatyzowane frazemy (gotowe w sytuacjach danego typu) i dopiero kontrola mówiącego dokonuje analizy stopnia odpowiedniości do relacjonowanego stanu rzeczy. Na podstawie zgromadzonych wcześniej praktyk językowo-sytuacyjnych, dochodzi do stanu antycypacji tego, co człowiek zamierza powiedzieć, gdyż samo zachowanie językowe w konkretnej sytuacji komunikacyjnej kreuje się jeszcze przed samym aktem wypowiedzi (Prangiszwili 1969: 88). Podobne stanowisko reprezentowali również badacze amerykańscy, którzy przedstawiali produkcję mowy jako odtwarzanie reakcji werbalnych, przyswojonych wcześniej drogą naśladowania innych użytkowników języka (Chlebda 2003: 92). Zaprezentowaną postawę frazematyka wpisuje w ciąg powtarzalności układu sytuacyjnego, który składać się może z komponentów „typ sytuacji → typowa reakcja werbalna”, przy czym zaistnienie znanego sygnalizatora typowości wywołuje typowe zachowanie werbalne. Dlatego też analizy frazematyczne rozpatrują pojęcie sytuacyjności w dwóch aspektach jako:

1. sferę, w której zanurzony jest każdorazowy akt komunikacji,
2. powtarzalność sytuacji językowych, ograniczająca człowieka w kwestii doboru środków językowych.

Układ sytuacyjny warunkuje dobór środków językowych wypowiedzi i wpływa na jej tematykę, cel, środek czy formę przekazu. Widoczny fakt determinacji określonego czynnika na szereg innych zjawisk narzuca obserwację aktu ludzkiej komunikacji z perspektywy repertuaru ograniczników swobody wypowiedzi. Innymi słowy, człowiek i jego język nie mogą istnieć jako byty w pełni niezależne. Według Chlebdy, przez wspomniane ograniczniki należy rozumieć „wszelkie ograniczenia, nakładające się zarówno na ludzkie zdolności poznawcze względem rzeczywistości, jak i na zdolność człowieka do możliwie adekwatnego odzwierciedlania jej elementów w formowanych przy pomocy języka sądach; tym samym ograniczenia owe nakładają się też pośrednio na działania w kategoriach pojęć skodyfikowanych w języku” (Chlebda 2003: 61). Do wspomnianych ograniczeń, których udział

występuje w różnych proporcjach, odnieść należy czynniki zarówno ekstralingwalne, jak i intralingwalne:

1. Ograniczniki ekstralingwalne: a) biologiczne, b) kulturowe, c) mentalne, d) sytuacyjne, e) stylowo-gatunkowe,

2. Ograniczniki intralingwalne.

Zawężając pole obserwacji tylko do ograniczników sytuacyjnych i ich wpływu na kształtowanie i dobór komponentów w wypowiedzi, uwzględnić należy fundamentalną funkcję języka, jaką jest każdorazowe *zwracanie się do kogoś*. Rola społeczna i sytuacja dodatkowo implikuje stosowanie przez mówiącego parametrów zmiennych, tj. tematyki wypowiedzi, środków i formy przekazu, których nieprzestrzeganie może naruszyć sprawność i spójność komunikacji. Forma *zwracania się do kogoś* jest uzależniona od znalezienia się osoby mówiącej w określonej społeczności językowej i sytuacji, gdzie wymagana jest zarówno znajomość repertuaru językowego danej grupy, jak i typowych zachowań werbalnych w odpowiedzi na komunikaty. Niektóre wydarzenia wymagają od mówiącego respektowania ścisłych rygorów językowych (np. udział w sytuacjach liturgicznych), które z jednej strony uniemożliwiają wprowadzenie do sytuacji twórczego komunikatu o charakterze indywidualnym, a z drugiej zapewniają stałość i powtarzalność użycia pewnych formuł wyrazowych. Język jest tworem uwarunkowanym przez tło pozajęzykowe, w którym zanurzony jest cały proces aktu mowy. Odwołując się do myśli Krystyny Pisarkowej i Wacława Cockiewicza, Wojciech Chlebda przyjął, że każdorazowy akt komunikacji stanowi wypadkową jednoczesnej interakcji następujących czynników:

1. Czynniki merytoryczne (kontekst treściowy):

- a) rodzaj tematu,
- b) stopień ustalenia tematu,
- c) sposób opracowania tematu.

2. Czynniki socjalne (kontekst społeczny):

- a) ranga i podział ról społecznych partnerów,
- b) wiek partnerów,
- c) pochodzenie partnerów,
- d) wykształcenie partnerów,
- e) uprzednia wiedza o przedmiocie wymiany zdań i o partnerze,
- f) stopień otwartości kontaktu i nastawienie mówiącego wobec rozmówcy.

3. Czynniki formalne (kontekst fizyczny):

- a) liczba uczestników aktu komunikacji,

- b) medium komunikacji,
- c) pora aktu komunikacji i inne jego parametry temporalne,
- d) miejsce aktu komunikacji,
- e) udział otoczenia fizycznego w akcie komunikacji i charakter tego otoczenia,
- f) obecność/nieobecność świadków aktu komunikacji (Chlebda 2003: 83–84).

Wszystkie wymienione wyżej aspekty wpisują się w charakterystykę kontekstu sytuacyjnego i tego, w jaki sposób i z użyciem jakich środków językowych będzie przebiegał akt komunikacji. Implikują one przymus zastosowania określonego gatunku mowy, dlatego też nie istnieją zupełnie dowolne w swym kształcie językowym wypowiedzi – składają się one z repertuaru formuł językowych, które w koncepcji Milniana Parry’ego określane są jako „grupa słów używana regularnie w tożsamy warunkach metrycznych dla wyrażenia określonego pojęcia” (Chlebda 2003: 88). W taki sposób frazematyka zanurza w sytuacji indywidualną wypowiedź mówiącego, która do tej pory była rozpatrywana w izolacji tekstowej i dowodzi, że znaczenie określonej jednostki odtwarzalnej zawsze jest uzależnione od szerszego kontekstu sytuacyjnego. Za egemplifikację przedstawionej tezy można przyjąć dotychczasowe wyjaśnienie w opracowaniach leksykograficznych zwrotu *daj spokój*, który był notowany tylko w znaczeniu „mówiący daje do zrozumienia nadawcy, żeby zaprzestał wykonywania określonej czynności”. Przedstawioną eksplikację należy naturalnie uznać za poprawną, aczkolwiek fakt pominięcia innych znaczeń sytuacyjnych owej jednostki unaocznia konieczność uwzględnienia kryterium sytuacyjnego i umieszczenia w słownikach pozostałych znaczeń i ilustracji, np.:

- daj spokój* (w odpowiedzi na czyjeś podziękowania),
- daj spokój* (prośba o zaprzestanie kontynuowania komunikatu),
- daj spokój* (wyrażenie zdenerwowania) itp.

Istotność uwzględnienia danego kryterium dla dalszych badań pozostaje niezaprzeczalna, co potwierdzają chociażby obserwacje Witolda Doroszewskiego, twierdzącego, że „w treści znaczeniowej każdego wyrazu tkwią elementy **sytuacji**, w których używaliśmy tego wyrazu lub spotkaliśmy się z jego użyciem, a wyobrażenia łączy się w impulsy, bodaj załączkowe do pewnych działań (Doroszewski 1970: 144, 156). Założoną we frazematyce gotowość wielowyrzawowych jednostek do ich użycia w komunikacji należy rozpatrywać jako gotowość względną, gdyż dopiero użycie wyrażenia w kontekście nadaje mu miano frazemu – produktu pełnego, wypełnionego formą i treścią znaczeniową.

2.5 Osoba mówiąca i model aktu komunikacji

Jak już zostało wspomniane, wszelkie rozważania frazematyczne skupiają się na analizie osoby mówiącej, która poprzez zwracanie się do kogoś wypełnia założone działanie i określony cel nawiązania kontaktu. Za rzecz istotną dla danego zagadnienia należy uznać zjawisko zajścia podwójnej adekwatności, która stanowi wymóg konieczny dla prawidłowej realizacji aktu komunikacyjnego, gdyż antycypacja tego, co zamierzamy powiedzieć, nie zawsze jest w pełni zgodna ze zdaniem wypowiedzianym.

Adekwatność pierwotna zakłada wysoki stopień tożsamości realizacji planu wypowiedzi do wcześniej powstałego zamysłu, a jej prawidłowy przebieg jest kontrolowany przez samego mówiącego.

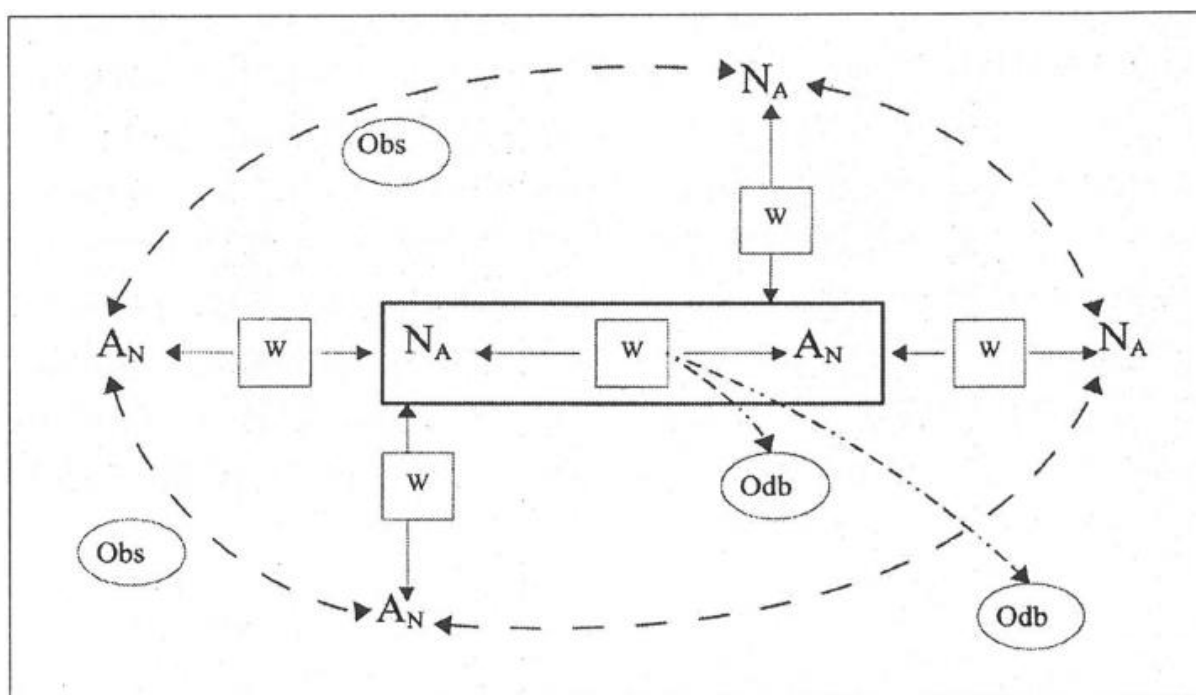
Z kolei przez *adekwatność wtórną* rozumie się zgodność między sensem komunikatu nadanego i przez rozmówcę odebranego.

Wojciech Chlebda wyróżnia dodatkowo dwa rodzaje tej adekwatności:

1. Między sensem komunikatu odebranego i sensem wynikającym z powierzchniowej warstwy komunikatu nadanego (*wtórna adekwatność bezpośrednia*),
2. Między sensem komunikatu odebranego i sensami głębokimi – zamysłem w umyśle nadawcy (*wtórna adekwatność pośrednia*) (Chlebda 2003: 109).

Frazematyka, rozpatrywana jako proces tworzenia indywidualnej wypowiedzi skierowanej do rozmówcy powinna więc uwzględniać schemat aktu komunikacji (np. uczestników, mechanizm przekazywania informacji) i semantyczną zawartość przekazywanych komunikatów. Semantyka frazemu oznacza więc znaczenie frazeologizmu zastosowanego w wypowiedzi przez określonego nadawcę w konkretnej sytuacji i z określoną intencją do drugiej osoby. Niezwykle istotne dla analizy osoby mówiącej pozostaje również obecność tzw. *samoświadomości językowej*, czyli „zdolności do uświadamiania sobie samego siebie jako „ja mówiącego”, proces uprzytomniania sobie podmiotowości swego mówienia oraz rezultat tego procesu” (Chlebda 2010b: 20). Frazematyka po raz pierwszy zwraca uwagę na umiejętność gromadzenia i przetwarzania informacji, którą w procesie wypowiedzi tworzy nadawca komunikatu i zarazem pozostaje jej pierwszym odbiorcą i słuchaczem. Osoba mówiąca dysponuje nie tylko możliwością natychmiastowej korekty swojej wypowiedzi, ale pozostaje zobowiązana do stałej kontroli wypowiadanego komunikatu, którego nieścisłości formalno-znaczeniowe mogłyby zaburzyć przebieg skutecznej komunikacji. Dlatego też samoświadomość językową uznać należy za wartość zmienną, zależną od indywidualnych kompetencji nadawcy i rozwijającą się drogą gromadzenia metainformacji językowej.

Frazematyczny model aktu komunikacji stoi w opozycji do stworzonych dotychczas propozycji (np. modelu komunikacyjnego Jakobsona), które dokonywały rozdzielania nadawcy i odbiorcy, nie uwzględniając ich wzajemnego aktywnego oddziaływania podczas kontaktu. Nadawca staje się jednocześnie odbiorcą: z jednej strony słucha samego siebie podczas formułowania wypowiedzi i kontroluje jej przebieg, a z drugiej antycypuje reakcję adresata, dostosowując do niej dobór środków językowych, aby wywołać pożądaną odpowiedź. Na grunt frazeologii zostaje więc wprowadzony zupełnie nowy schemat przebiegu komunikacji, uwzględniający nowe elementy, które stają się wyznacznikami porozumiewania się we frazematyce (Chlebda 2003: 115):



Nadawca, formułując swą wypowiedź pozostaje odbiorcą wcześniejszych wypowiedzi innych osób, przez co najczęściej wykorzystuje zasłyszane jednostki do prowadzenia rozmowy. Oznaczony na schemacie nadawca (N_A), który bywa adresatem i związane jest z nim pojęcie zamierzonej intencji w stosunku do konkretnego adresata. Innowacyjnym elementem tego modelu jest wprowadzenie kategorii odbiorcy przypadkowego, rozumianego jako osoba, do której komunikat nie był adresowany, lecz która w określonych warunkach komunikacyjnych staje się jego niezamierzonym odbiorcą. Dlatego też każdy adresat jest odbiorcą wypowiedzi, lecz nie każdy jej odbiorca jest adresatem. Istotnym również jest umieszczenie w danym modelu roli obserwatora (Obs), który z dystansu (również czasoprzestrzennego) obserwuje przebieg aktu komunikacji i nie bierze czynnego udziału w prowadzonym dialogu. Dysponuje określoną wiedzą językoznawczą, ocenia elementy, składające się na przebieg rozmowy i ma

możliwość postrzegania wypowiedzi w szerszych zależnościach komunikacyjnych. Należy również wspomnieć, że grubsza ramka na ilustracji symbolizuje elementarny łańcuch komunikacji, a linia przerywana zarysowuje granice dyskursu i intertekstu. Z kolei symbol W oznacza tekst wypowiedzi, funkcjonujący w trzech aspektach:

- w fazie emisji,
- w fazie obiektywnego bytu,
- w fazie odbioru (Gajda 1988: 24).

Najistotniejszą dla naszych rozważań jest faza druga, która umożliwia postrzeganie wypowiedzi bez odnoszenia jej do konkretnego nadawcy i odbiorcy, stwarzając tym samym szansę wyodrębniania określonych schematów komunikacyjnych i składających się na nie komponentów (frazemów).

2.6 Status semantyczny frazemów

W tradycyjnej frazeologii zakładano, iż jednostki frazeologiczne w większości charakteryzują się jednoznacznością znaczenia, co wynika z ich umiejscowienia w konkretnym kontekście, pozbawiając ich tym samym możliwości dodatkowych interpretacji. Przy zastosowaniu wyrazu polisemicznego kontekst miał pełnić funkcję selekcyjno-determinacyjną, a przypadki zachowania przez jednostkę wieloznaczności w wypowiedzi uznawano za przypadki pojedyncze, niepodlegające dokładniejszej analizie. W ujęciu frazeologii pragmatycznej frazemy (i węższej frazeologizmy) cechują się naturą dwuprzestrzenną znaczeniowo, a w ich statusie semantycznym zakodowane są różne możliwości ich interpretacji, założonych intencji i odbioru. Warunkiem koniecznym do udziału w dyskursie jest umiejętność dostrzeżenia podwójnej modalności informacji, polegającej na odniesieniu wypowiedzi do rzeczywistości niejęzykowej i jej prawidłowym odczytaniu z perspektywy intencji nadawcy. Frazem jako nieodłączny element procesu tworzenia i odbioru wypowiedzi stanowi jednocześnie komponent gotowego bloku wyrazowego, jak i zachowuje potencję oddzielności tych komponentów, dążąc do syntagmatycznego dopełnienia innym znakiem. Niemniej jednak, częstą przyczyną nieporozumień w komunikacji może być fakt, że między dwoma uczestnikami dialogu w pełni zrozumiały jest jedynie wspólny kod językowy, natomiast sens wypowiedzi i jego odbiór uzależnione są od indywidualnych doświadczeń i świadomości rozmówców. Frazematyka traktuje znaczenie leksykalne jako rodzaj fenomenu psychicznego, gdyż rzeczywista semantyka komunikatu jest w pełni znana tylko jej nadawcy i nie musi być zbieżna ze znaczeniem zrozumiałym przez odbiorcę. Istotną rolę w tym procesie odgrywa

uprzednia wiedza osób, dotycząca spodziewanych w rozmowie typów znaczeń, wśród których wyróżnić można znaczenia podstawowe i symboliczne. Wiedza o apriorycznym prawdopodobieństwie rozkładu znaczeń i dwuprzestrzenności znaczeniowej określonych pojęć pozwala uniknąć nieścisłości komunikacyjnych, jednak nie zawsze jest to możliwe ze względu na odmienny stopień poziomu semiotycznego rozmówcy. W takim rozumieniu przedstawić należy cztery ogólne typy sytuacji semantycznych:

1. Nazwa w kontekście jest dla Nadawcy jednoznaczna i dla Odbiorcy jednoznaczna;
2. Nazwa w kontekście jest dla Nadawcy jednoznaczna, dla Odbiorcy wieloznaczna;
3. Nazwa w kontekście jest dla N wieloznaczna, dla O jednoznaczna;
4. Nazwa w kontekście jest dla N wieloznaczna i dla O wieloznaczna (Chlebda 2003: 134).

Dla skuteczności aktu komunikacji zasadniczy jest odpowiedni dobór przez nadawcę takich środków, które umożliwiłyby odbiorcy aktualizację zamierzonego znaczenia oraz właściwe rozpoznanie relacji między poszczególnymi elementami komunikatu. Frazematyka nie skupia się wyłącznie na traktowaniu kontekstu jako jedynego słusznego aktualizatora, lecz podkreśla możliwość oddziaływania kilku różnorodnych aktualizatorów w tym samym czasie w wypowiedzi. Wspomniane aktualizatory znaczeń rozdzielić można na następujące grupy: 1. Językowe aktualizatory proste: a) kontaktowe, b) dystansowe; 2. Językowe aktualizatory pośrednie: a) Formuły wskaźnikowe dosłowności, b) formuły wskaźnikowe przenośności, c) Formuły dwuprzestrzenności znaczeniowej, d) wskaźnikowe formuły opisowe; 3. Aktualizatory parajęzykowe: a) graficzne, b) prozodyczne, c) kinetyczne, 4. Aktualizatory parajęzykowe. Dla egzemplifikacji jednego z nich może posłużyć następujący przykład:

Śmiem twierdzić, że w całym okresie transformacji, choć budżety były bardzo różne, wszystkie były nastawione nie na rozwój, nie na to, żeby wspierać przedsiębiorczość, wspierać i promować polski eksport, ale były nastawione na stagnację, na przetrwanie i gaszenie pożaru. Tym budżetem znów staramy się ugasić pożar – **mówiąc w przenośni** – który zgotowali nam kolejarze, a cały czas gdzieś tam strajkują też rolnicy.

W danym tekście uwidacznia się użycie wskaźnikowej formuły przenośności, która wyklucza możliwość literalnego odczytania wyrażenia i narzuca symboliczne odczytanie „zgotowania pożaru” przez kolejarzy. Niemniej jednak status językowy danego frazemu uwzględnia jego dwuprzestrzenność znaczeniową i w innym otoczeniu kontekstowym mógłby być interpretowany w dosłownym znaczeniu, tj. „wyraz po wyrazie”. Przedstawione powyżej obserwacje skłaniają do wniosku, że w tekstach możliwe jest zaistnienie realnej bądź potencjalnej dwuprzestrzenności znaczeniowej, co stoi w opozycji do dotychczasowego postrzegania znaczeń w ujęciu semantyki: „Za udowodnione i nie podlegające wątpliwości

uważa się dzisiaj twierdzenia, że wyraz może posiadać dwa lub kilka znaczeń w konkretnym tekście lub w konkretnej sytuacji dany wyraz zrealizuje zawsze tylko jedno znaczenie” (Kominicz 1971: 42–43). We frazematyce wspomniana wieloznaczność jest regularną cechą każdej wypowiedzi, a każdy frazem umieszczony jest w siatce powiązań i oddziaływań znaczeniowych, będących wynikiem komunikacji między nadawcą a odbiorcą. Uwzględnienie aspektu dwuprzestrzenności pozwala podzielić frazemy ze względu na ich status semantyczny na następujące typy:

1. **Frazem wirtualny** (o charakterze inwariantu, posiadający wszystkie znaczenia naraz bez rozpatrzenia osobnych użyć w kontekście).
2. **Wariant frazemu** – jedno ze znaczeń typowych, będących składnikiem inwariantu.
3. **Frazem aktualny** – postać konkretnego wykorzystania frazemu w wypowiedzi (Chlebda 2003: 142).

Tradycyjna frazeologia wraz z opracowaniami leksykograficznymi skupiała się na notowaniu przykładów frazemów inwariantowych, stanowiących typowy przykład jednostki słownikowej o znaczeniu możliwie jak najbardziej ogólnym. Tożsame im frazemy wirtualne mają wpisany w swoją naturę status dwuprzestrzenny, gdyż uwzględniają wszystkie możliwe znaczenia, zarówno podstawowe (dosłowne) jak i symboliczne (metaforyczne). Warianty frazemu stanowią część składową postaci inwariantowej, odnoszą się tylko do jednej z wymienionych płaszczyzn, a ich jak najbardziej reprezentatywny zbiór umożliwia stworzenie wyczerpującego opisu słownikowego dla leksykografii jedno- i dwujęzycznej. Z kolei kształt frazemu aktualnego zależy od indywidualnych zamierzeń nadawcy i odbiorcy tekstu, a organizacja i rozkład aktualizatorów przyczyniają się do szerokiego repertuaru dopuszczalnych interpretacji semantycznych wypowiedzi. Za przykład może posłużyć frazem *koń trojański*, którego etymologię naturalnie należy łączyć z mitologiczną sytuacją wojowników greckich podczas zdobywania Troi. Od frazemu inwariantowego można przejść do jednego ze znaczeń, które pod danym pojęciem rozumie również złośliwe oprogramowanie komputerowe, na wzór swojego pierwowzoru podszywające się pod przydatne dla systemu aplikacje. Z kolei za ilustrację dla frazemu aktualnego może posłużyć użycie tej jednostki dla określenia szkodliwego działania osoby na rzecz organizacji, której jest członkiem:

W meczu Włochy - Polska (30 kwietnia) w Neapolu trener kadry nie wystawił zawodnika do gry. Piłkarz nazwał metody pracy szkoleniowca archaicznymi.
- Dla Kowalczyka nie ma już miejsca w moim zespole. To był *koń trojański* reprezentacji.

Współczesny opis leksykograficzny frazemów powinien uwzględniać zarówno warstwę znaczeń podstawowych, jak i szereg znaczeń przenośnych w ilustracjach kontekstowych, co umożliwi odzwierciedlenie realnego statusu językowego frazemów i ocenę wypełnianych przez nie funkcji w tekstach. Skrupulatna kodyfikacja słownikowa frazemów, uwzględniająca cechę odtwarzalności względnej, zawierać winna ich zróżnicowanie w postaci wielowariantowej i wielokontekstowej. Z kolei ścisła parametryzacja opisu przyczyni się do dokładniejszego zakresu możliwości jednostki (charakterystyki semantyczne, składniowe, stylistyczne czy pragmatyczne), aczkolwiek nie jest możliwe przewidzenie wszystkich jej realizacji tekstowych. Przyjęta w słowniku forma inwariantu określonej jednostki wydaje się być z jednej strony zabiegiem koniecznym do ustalenia znaczenia uogólnionego, a z drugiej staje się jej ukonkretnieniem, pozbawionym istoty abstrakcyjności. Przedstawione rozważania skłaniają do wniosku, iż frazematyka nie dokonuje rozgraniczenia pomiędzy pojęciami *znaczenie* a *użycie*, a traktuje je jako byty wzajemnie równoważne, uwzględniając naturalną zmienność znaczenia w zależności od rozpatrywania wypowiedzi nadawcy bądź odbiorcy. Ten sam komunikat może być diametralnie różnie rozumiany przez uczestników aktu komunikacji, na co wpływa pierwotna dwuprzestrzenność znaczeniowa frazemów, sprowadzająca się do jak odczytywania analitycznego (dosłownego), tak i syntetycznego (ponadwyrazowego, symbolicznego). Niezrozumienie w komunikacji jest jej częścią składową, a proces wymiany myśli między nadawcą a odbiorcą zawsze oscyluje pomiędzy biegunem pełnego i niepełnego porozumienia, wynikającym z odmiennej wiedzy o świecie uczestników komunikacji. Potencjalna obecność dodatkowej interpretacji znaczeniowej wypowiedzi otwiera możliwości dla napędzenia dyskusji, ponadplanowej wymiany zdań i ruchu myśli w kierunku całościowego zrozumienia pojęć. Zastosowanie w wypowiedzi wspomnianych aktualizatorów pozwala uniknąć nieporozumień w odbiorze informacji, dzięki czemu aktualna wieloznaczność frazemu przekształca się w zrealizowaną potencję i naprowadza odbiorcę na prawidłowe odcodowanie znaczeń. W takim ujęciu „gotowość” jednostki frazeologicznej na tle paradygmatów frazematycznych należy rozpatrywać jako gotowością względną, która dopiero w określonym kontekście i po konkretnych przekształceniach nadawcy jest zdolna zyskać status produktu pełnego, uwzględniającego zarówno formę, jak i treść frazemu.

2.7 Reprodukowalność we frazematyce

Na gruncie badań językoznawczych pojęcie odtwarzalności przeważnie było sytuowane w kręgu paradygmatów frazeologicznych, pomimo wyraźnych rozbieżności frazeologów

dotyczących kwestii uznania określonych połączeń wyrazowych za reprodukowalne. W odróżnieniu od założeń frazematyki owa odtwarzalność stanowi w tradycyjnym ujęciu frazeologii czynnik wyróżniający jednostki frazeologiczne, nie będąc ich cechą konstytutywną. Odtwarzalność we frazematyce nie tylko jest jej cechą fundamentalną, ale jednocześnie zawiera w sobie zdolność wielokrotnego występowania w tekstach wielowyrazowca, używanego przez różnych nadawców i w odmiennej przestrzeni czasowej, którego postać graficzna i foniczna względnie jest stała. Uprzednio ukształtowany w pamięci językowej zespół treściowy i jego kształt umożliwia zwerbalizowanie i odpowiednie formułowanie myśli. Odtwarzalność według Szanskiego, łączy się z gotową postacią składu leksykalnego frazeologizmów, których poszczególne elementy znajdują się w niezmiennym i ściśle ustalonym porządku, przybierając kształt gotowych jednostek języka (Шанский 1996: 26–28). Bliższe frazematyce są założenia Grigorija Piermiakowa, zaliczającego do repertuaru odtwarzalnych klisz językowych nie tylko frazeologizmy, ale i idiomy, porzekadła, przysłowia, aforyzmy, skrzydlate słowa itp., czyli dokładnie te jednostki, które można opatrzyć hiperonimem *frazemy* (Пермяков 1970). Wymienione obserwacje skłaniają do wysunięcia wniosku, iż założona „gotowość i niezmienność” jednostek frazeologicznych w mowie i piśmie w rzeczywistości narzuca ich wariantywność składu leksykalnego i syntaktycznego.

Dotychczasowe dążenia do ustalenia inwariantu jednostki, który odzwierciedlałby jej wszelkie realizacje tekstowe jest wyraźnym uproszczeniem, gdyż każdy wariant ze swoim indywidualnym znaczeniem kontekstowym pozostaje w ścisłej korelacji z inwariantem, wpływając na jego treść i formę. Wymienność i przekształcenia w strukturach frazemów dotyczą zarówno składników formalnych, jak i wypełnienia leksykalnego komponentów. Z tego względu każdorazowy opis wariantu powinien zawierać jego charakterystyki stylowe, ekspresyjne czy parametry łączliwościowe. Walerij Mokijenko zwraca uwagę na fakt, iż źródło każdego frazeologizmu pochodzi ze swobodnego połączenia wyrazów o wysokim stopniu wymienialności komponentów, wskutek czego powstaje ruch od wariantowej różnorodności jednostek do jednorodności składniowo-leksykalnej (Мокиенко 1980: 13). We frazematyce każdorazowe użycie frazemu i jego realizacja tekstowa prowadzi do wytworzenia nowego wariantu, naruszając minimalnie tym samym strukturę wzorca słownikowego i nadając mu charakter gotowości/stałości względnej, a nie absolutnej. Niezbędne więc do wyodrębnienia formy inwariantowej jest odnalezienie wspólnego semantycznego mianownika dla jego możliwych wariantów, cechujących się podobną strukturą składniową i brakiem naruszenia „semantycznej stabilności, regulowanej w określonym stopniu przez stabilność struktury” (Мокиенко 1980: 38).

Za główne tendencje języka, pozostające w nieustannym sprzężeniu zwrotnym, uznaje się proces zastygania określonego składu leksykalnego i jego jednoczesnego odnowiania, które pozwala dostosować treść i strukturę frazemu do konkretnych potrzeb komunikacyjnych. Odtwarzalność we frazematyce rozumiana będzie więc jako *odtwarzalność względna*, tylekroć zmieniająca układ i zawartość połączeń wyrazowych, ilekroć wymaga tego sytuacja i użycie przed indywidualnego użytkownika języka. Dlatego też postać wyrazową frazemów należy uznawać za wielkość nieustabilizowaną, znajdującą się w procesie nieprzerwanego kształtowania się i wariantowości elementów, pozwalającą na przystosowanie się wyrażenia do nowszych jakościowo kontekstów i sytuacji. W odróżnieniu od zasad frazeologii następuje tutaj przypisanie cech odtwarzalności wszystkim „stabilnym” połączeniom wyrazowym (np. *ładny stół; przystanek autobusowy; a to ci heca!*), służącym do nominacji desygnatu czy ocenie sytuacji. Wspomniany wcześniej trychotomiczny ciąg Olgi Dołgowej: „odtworzenie czyjegoś tworu mowy – stworzenie tworu nieoryginalnego – stworzenie tworu oryginalnego” wpisuje się we frazematyczne odrzucenie dychotomii *produkcji – reprodukcji* i *twórczości – odtwórczości*. Ponadto umiejscowienie w akcie odtwórczym pojęcia intencji podkreśla każdorazowy charakter celowości wypowiedzi, a zaistnienie „intencji mownej” przyczynia się do ustanowienia nadawcy, jego nastawienia i określonej sytuacji komunikacyjnej.

Innym problemem badawczym w ramach przedstawionego paradygmatu frazeologii nadawcy pozostaje odtwarzalność frazemów nie tylko „beziemiennych” i stanowiących dobro powszechne ogółu społeczności, ale również mających swojego autora czy bezpośrednio źródło pochodzenia. Dotychczasowe ujęcie frazeologizmów jako jednostek wyłącznie „niepodpisanych” wyklucza z kręgu frazeologii cechujące się reprodukowalnością i „gotowością” powtórzenia, przywołania, aluzje, przytoczenia czy cytaty, w przypadku których stopień świadomości ich użycia przez użytkownika nie jest dla frazematyka kwestią nadrzędną dla ich klasyfikacji. Wszystkie te zjawiska można sprowadzić do zbioru frazemów, zawierającego różnorodne formalnie, gatunkowo i znaczeniowo ciągi wyrazowe. Głębszej analizie natomiast poddawane zostają wszelkie przekształcenia w wyniku niedokładnego odwzorowania wyrażenia, pociągające ze sobą zmiany strukturalne, stylistyczne i semantyczne i tym samym przyczyniają się do stworzenia nowego frazemu. Spełnienie przez jednostkę funkcji nominacyjnej do opisu pojedynczych pojęć i podstawowych potencjałów treściowych, przy zachowaniu swej względnej odtwarzalności, każdorazowo czyni z niej jednostkę frazematyczną, gotową do użycia w swej konkretnej postaci. Łączona z reprodukowalnością powtarzalność frazemów, stopień frekwencji występowania w określonych tekstach i zachowanie ich funkcji komunikacyjno-poznawczej jest zarazem

podstawą dla ich wyodrębniania. Zastosowanie analizy statystycznej frazemów, uwzględniającej czynniki komunikacyjne, nawiązuje pośrednio do myśli Emanuela Szubina, który zalecał rozpoczęcie analizy znaków rekursywnych (powtarzalnych) w tekstach od zweryfikowania jednostek najbardziej złożonych, a nie pojedynczych wyrazów. Wnikliwa obserwacja tekstów i powtarzających się ciągów wyrazowych umożliwia wyodrębnienie frazemów, stanowiących już część pamięci językowej i tym samym zakwalifikowanie jednostki jako reprodukowalnej czy frazematycznej.

Na zakończenie naszych rozważań nie sposób nie wspomnieć o odejściu Wojciecha Chlebdy od terminu *frazem* i sformułowaniu nowego pojęcia, uwzględniającego w swej nazwie priorytetowe zjawisko reprodukowalności, czyli *reproduktu*. Przez dane określenie należy rozumieć „jednostką języka, wyodrębnioną z tekstów, sformułowanych w tym języku, w rezultacie stwierdzenia jej regularnej powtarzalności w tych tekstach w funkcji werbalizatora tego, co autor chciał powiedzieć, czyli pewnej wiązki sensów, intencji, emocji, określonego zespołu treściowego” (Chlebda 2010a: 15). Stosowana we frazeologii nadawcy definicja frazemu uwzględnia jego zastosowanie w określonych sytuacjach komunikacyjnych dla wyrażenia konkretnego potencjału treściowego, co czyni go odtwarzalnym i wpisuje go w system frazeologiczny danego języka. Z kolei reprodukt uwydatnia w swej nazwie priorytetową cechę odtwarzalności, traktowanej jako powtarzalność wielowyrazowego ciągu w tekstach wypowiedzi ludzkich. Stworzenie bliskoznacznego i nowego terminu było spowodowane zaistniałym w lingwistyce chaosem terminologicznym, ponieważ termin *frazem* był używany w innych koncepcjach lingwistycznych (Lewicki, Mielczuk, Awdiejew) i służył opisowi innych zjawisk. Reprodukt nie tylko wchodzi w opozycję z pojęciem produktu języka, ale jako nośnik rzeczonych pojęć ma pobudzać związki asocjacyjne między odtwarzalnością a oznaczonym znakiem językowym. Ukształtowane przez tradycję językowo-kulturową reprodukty służyć więc będą nazywaniu fragmentów rzeczywistości, pojęć czy zamysłów, których użycie w sytuacji można uznać za zachowanie odruchowe. W pewnym stopniu nowe pojęcie pokrywa się ze wspomnianą jednostką języka w ujęciu Andrzeja Bogusławskiego i nominatem Grażyny Sawickiej, przez który badaczka rozumiała każdą samodzielną jednostkę języka, charakteryzującą się stałością związku formy i treści oraz denotacją, przy czym forma nominatu może przybrać formę jedno- lub wielowyrazową (Sawicka 1995: 15). Reprodukt różni się od przedstawionego opisu jedynie wprowadzeniem do obszaru swoich badań również całego zdania.

ROZDZIAŁ III

Lingwistyka Korpusowa. Rys historyczny

Początki nowego działu językoznawstwa należy datować na lata pięćdziesiąte XX wieku, a jego powstanie łączy się z silną reakcją na postulaty gramatyki transformacyjno-generatywnej Noama Chomsky'ego. W jego rozumieniu intuicja każdego rodzimego użytkownika języka w kwestii orzeczenia poprawności gramatycznej zdań jest wartością niezaprzeczalną, wynikającą z wrodzonej znajomości struktur i reguł gramatycznych (Wołkowski 2010: 148). Koncepcja ta wykluczała więc niekwestionowaną różnorodność opinii i sądów użytkowników języka, dotyczących określonej struktury zdaniowej, zawężyła kreatywność językową do produkcji nowych okazów (a nie typów) i doprowadziła do oddzielenia w języku zdań dosłownych od przenośnych figur mowy. W opozycji powstał nurt odmienny, wywodzący się od Williama Labowa, który uczynił uzus językowy płaszczyzną dla przeprowadzenia wiarygodnych badań w oparciu o częstość występowania poszczególnych form językowych (Bell, Sharma, Britain 2016: 399). Funkcjonalne modele użycia języka i określonych jednostek wiązało się z pojęciem prawdopodobieństwa wystąpienia konkretnej formy w różnych odmianach języka, a jednoczesny wzmożony rozwój technik komputerowych przyczynił się do możliwości przechowywania ogromnej ilości zbiorów tekstowych na nośniku elektronicznym, dając tym samym początek językoznawstwu korpusowemu.

Językoznawstwo korpusowe to dział lingwistyki komputerowej, zajmujący się wykorzystaniem technik informatycznych w celu stworzenia dużego, ujednoliconego, adnotowanego, reprezentatywnego korpusu językowego, który umożliwia analizę określonych zjawisk lingwistycznych. Odpowiednio duża objętość korpusów zapewnia możliwość obserwacji całego spektrum zjawisk językowych, a występowanie jednostek w swoim naturalnym środowisku tekstowym pozwala na ich wieloaspektowe badania. Pierwszy korpus językowy autorstwa W. Francisa i H. Kucery powstał na Uniwersytecie Browna w USA i stanowił ponad półmilionowy zbiór tekstów amerykańskiej odmiany angielszczyzny, uwzględniając przy tym dużą liczbę słowników frekwencyjnych do opracowań statystycznych. Pojawienie się korpusu nie tylko wzbudziło powszechne zainteresowanie i ożywione dyskusje, ale przede wszystkim spowodowało gwałtowny rozwój korpusów w latach 80. XX wieku, o czym świadczy następujące stwierdzenie: «в среде ученых появилось понимание того, что ряд корректных лингвистических исследований возможно провести только на большом

речевом материале» (Мамонтова 2010: 232). Pod koniec lat 90 XX w. w krajach słowiańskich ruszyły intensywne prace nad własnymi korpusami narodowymi, w których nieunikniona konieczność uwzględnienia fleksji przyczyniła się do stworzenia szerokiego aparatu dla automatycznej analizy morfologiczno-składniowej czy rozróżniania przez narzędzie zjawiska homonimii form gramatycznych. Współcześnie najbardziej reprezentatywnym i najstarszym korpusem słowiańskim pozostaje Narodowy Korpus Języka Czeskiego, w którego zasobach znajdziemy ponadto oddzielne bazy tekstowe dla ponad sześćdziesięciu języków.

W tym samym czasie w Rosji lingwistyka korpusowa zaczęła się rozwijać, tworząc najstarsze korpusy rosyjskojęzyczne, tj. Успальский корпус русских текстов, składający się z 600 fragmentów literatury pięknej i tekstów informacyjnych, a także powszechnie udostępniony Тюбингенский корпус (Шаров 2003: 11). Narzędzia wyposażono w system znaczników anotacji morfologicznych, co umożliwiło wyekscerpowanie tekstu wraz z gotowymi oznaczeniami. Przed powstaniem tychże baz tekstowych możemy mówić o sporządzeniu korpusu prototypowego, którym był Частотный словарь русского языка, zbudowany na podstawie prostych plików tekstowych o objętości miliona wyrazów i uwzględniający 4 gatunki tekstowe. Jednym z kolejnych przedsięwzięć był Машинный Фонд русского языка, którego celem miało być stworzenie reprezentatywnego korpusu różnych gatunków i innowacyjnych technik cyfrowych, umożliwiających przetwarzanie języka naturalnego i opracowanie nowych słowników (Захаров 2014). Gromadzenie danych tekstowych, leksykograficznych i gramatycznych miało posłużyć analizom językoznawczym i wdrażaniu stosowanych rozwiązań. Jednak po roku 1991 prace nad utworzeniem korpusu zostały wstrzymane. Mimo to niektóre teksty korpusu z XIX wieku zostały oddane do użytku ogólnego. Teksty prasowe zawierał Компьютерный корпус текстов русских газет конца XX в., w którym dobór tekstów przeprowadzono w oparciu o zasady włączania do korpusu kompletnych wydań 13 gazet rosyjskich z lat 1994–1997 (np. „Литературная газета”, „Правда” itp.), co dało łączną liczbę 23 110 tekstów i pozwoliło uzyskać obiektywny obraz tekstu prasowego. W roku 2001 rozpoczęto prace nad bazą rosyjskich tekstów internetowych pod nazwą Хельсинкский аннотированный корпус русских текстов, stanowiący część projektu Składni Funkcjonalnej Języka Rosyjskiego, anotowany morfologiczno-syntaktycznie. Zbiór około 100 tysięcy użyć słów z czasopisma „Итоги” rejestruje stylistykę języka rosyjskiego wraz z szeroką gamą współczesnych gatunków dziennikarskich. Jednak za największe osiągnięcie lingwistyki korpusowej w Rosji należy uznać pojawienie się Narodowego Korpusu Języka Rosyjskiego (ros. Национальный корпус русского языка),

korpusu monitorującego², obecnie zawierającego 6 228 828 tekstów różnych gatunków. Reprezentatywność korpusu i ogromna baza danych tekstowych, a także rozbudowany system tagsetu³ czyni korpus ten najcenniejszym materiałem rosyjskim do badań ilościowych i jakościowych.

Rozwój korpusów w Polsce należy datować na początek lat dziewięćdziesiątych, kiedy to grupy polskich językoznawców, informatyków i leksykografów tworzyły podwaliny pod rozwój językoznawstwa korpusowego. Tworzeniem baz tekstowych i opracowaniem narzędzi do ich praktycznego stosowania, takich jak tagery czy parsery składniowe⁴, zajmowali się przede wszystkim Zygmunt Saloni, Marek Świdziński, Krzysztof Szafran czy Zygmunt Vetulani. Jedną z najstarszych grup lingwistyki korpusowej, działającej w Katedrze Języka Angielskiego na Uniwersytecie Łódzkim, pozostaje zespół PELCRA, który w ramach wspólnego projektu z Katedrą Językoznawstwa Uniwersytetu w Lancaster zgromadził zarówno polskie dane językowe, jak i oryginalne tłumaczenia angielsko-polskie i polsko-angielskie (Lewandowska-Tomaszczyk 2005: 19). Współczesne zasoby PELCRA zawierają Narodowy Korpus Języka Polskiego, którego zbiory danych są morfologicznie anotowane. Korpus ten zawiera ponad 130 mln wyrazów, różne typy tekstów i odmian polszczyzny, a także stanowi bazę materiałową dla szeregu innych badań, np. do tworzenia *Wielkiego Słownika Języka Polskiego* czy Korpusu Polsko-Rosyjskiego Uniwersytetu Warszawskiego. Istotną częścią korpusu PELCRA jest Polski Korpus Uczniowski Języka Angielskiego, liczący ponad 1,5 miliona słów dotyczących danych językowych Polaków, uczących się języka angielskiego na różnych poziomach. Innym z zespołów badawczych, który przyczynił się do stworzenia w latach 2001-2004 ówczesnie największego i znakowanego morfosyntaktycznie korpusu języka polskiego, jest Zespół Inżynierii Lingwistycznej IPI PAN. Aktualne prace skupiają się na projektach CESAR, ATLAS, NEKST i CLARIN, z których ostatni stanowi przykład rozbudowanego systemu narzędzi, potrzebnych do przetwarzania dużej liczby tekstów dla obranych celów badawczych (np. znakowanie fraz, poprawianie interpunkcji w tekście czy wydobywanie terminów z korpusów) (Lewandowska-Tomaszczyk i in. 2012: 6). W 1997 roku w trakcie prac nad pierwszym słownikiem języka polskiego opartym na korpusie powstał zespół projektowy PWN przy Redakcji Słowników Języka Polskiego. Na potrzeby leksykograficzne szczególnie uwzględniono bazę ilustracji kontekstowych opracowywanych haseł i ich

² Przez korpus monitorujący należy rozumieć otwartą bazę tekstową, regularnie aktualizowaną i uzupełnianą o nowe teksty.

³ Tagset jest zbiorem indeksów używanych w procesie anotowania korpusów językowych.

⁴ Parser - program komputerowy dokonujący analizy składniowej danych w celu określenia ich struktury gramatycznej.

różnorodność tematyczno-stylistyczną. Finalnie osiągnął on objętość ponad 100 milionów słów i stał się podstawą dla przygotowania *Innego słownika języka polskiego* (Bańko 2000) i *Uniwersalnego słownika języka polskiego* (Dubisz 2003).

Rozwój językoznawstwa korpusowego w Czechach łączyć należy przede wszystkim z osobą Františka Čermáka, który tworząc ponad dwadzieścia lat temu Instytut Narodowego Korpusu Języka Czeskiego (czes. *Ústav Českého národního korpusu*), położył podwaliny pod najlepszy słowiański projekt korpusowy. Początkowo skromne przedsięwzięcie, powstałe na Wydziale Filozoficznym Uniwersytetu Karola we współpracy z kilkoma instytucjami i uniwersytetami z biegiem czasu stało się ośrodkiem czeskiej lingwistyki korpusowej i główną bazą danych dla badań lingwistycznych. Powstała w 1991 roku inicjatywa *Počítačový fond češtiny*, była wyrazem zarówno zapotrzebowania na nowe źródła informacji, jak i niezadowolenia z ówczesnego stanu językoznawstwa czeskiego i braku wykorzystania w badaniach szerszych danych tekstowych (Čermák 2017: 69). Wcześniejsze doświadczenie zespołu z korpusami anglojęzycznymi pozwoliło skonstruować analogiczną bazę informacji z uwzględnieniem rozwiązań problemów natury technicznej, wynikających ze specyfiki języka czeskiego. Rozwój działalności ÚČNK został wówczas wsparty poprzez uzyskanie kilku dotacji od *Grantové agentury České republiky*, wielu ministerstw bądź sponsorów prywatnych, takich jak *Nakladatelství Lidové noviny*. W tworzeniu korpusu uczestniczy współcześnie szereg instytucji, do których należą m.in.: Instytut Języka Czeskiego i Teorii Komunikacji (*Ústav českého jazyka a teorie komunikace*), Instytut Studiów Bohemistycznych (*Ústav bohemistických studií*), Instytut Lingwistyki Teoretycznej i Komputerowej (*Ústav teoretické a počítačové lingvistiky*), Instytut Lingwistyki Formalnej i Stosowanej - Wydział Matematyczno-Fizyczny Uniwersytetu Karola (*Ústav formální a aplikované lingvistiky - Matematicko-fyzikální fakulta UK*), Instytut Języka Czeskiego - Wydział Filozoficzny Uniwersytetu Masaryka (*Ústav českého jazyka - Filozofická fakulta MU*).

Wraz z systematycznym budowaniem korpusów i rozwijaniem ich specjalistycznego aparatu metodologicznego, projekt ma również na celu udostępnianie szeregu danych do analiz badawczych, np. w obszarze automatycznego przetwarzania języka naturalnego czy opracowywania narzędzi do korpusów eksploracyjnych. Zasoby Narodowego Korpusu Języka Czeskiego pozostają ogólnodostępne i szeroko wykorzystywane zarówno przez lingwistów i ekspertów różnych dziedzin, jak i osoby, zainteresowane konkretnymi zagadnieniami językoznawczymi. Od stycznia 2014 r. możliwym stało się przeszukiwanie korpusów ÚČNK za pomocą menedżera KonText, stanowiącego podstawową wyszukiwarkę w dostępnych bazach tekstowych. Spośród korpusów czeskich odnaleźć można zarówno korpusy

synchroniczne języka czeskiego pisanego, korpusy mówione synchroniczne, jak i korpusy diachroniczne. Do największych z nich należą: korpusy serii SYN (synchroniczne korpusy pisane), korpus korespondencji prywatnych (*KSK-dopisy*), praski i brneński korpus języka mówionego czy seria ORAL (transkrypcje rozmów). Jednym z największych osiągnięć pozostaje stworzenie dużego równoległego korpusu synchronicznego InterCorp, uwzględniającego bazy tekstowe dla ponad sześćdziesięciu języków.

3.1 Korpus językowy

Upowszechnienie się językoznawstwa korpusowego w różnych zespołach badawczych, działających w odmiennych częściach świata przyczyniło się do wytworzenia wielu definicji korpusów, które w większości zawierają jednak elementy wspólne. Przed przejściem do dalszych rozważań warto przytoczyć niektóre z nich:

A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research (Sinclair 2004) [dostęp 17.12.2023].

Корпус некоторого языка – это, в первом приближении, собрание текстов на данном языке, представленное в электронной форме и снабженное научным аппаратом (Плунгян 2005: 6).

Под названием языковой корпус понимается большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач (Захаров 2005: 3).

Ve smyslu dnešním se v korpusové lingvistice ale korpusem rozumí především takový ucelený soubor textů, který je sestavený s přihlédnutím ke svému cíli reprezentativním způsobem, je obhospodařovaný počítačové a zpracováván souborem korpusových metod. Častým předpokladem pro takové zpracování je anotace korpusu, umožňující získat nejen informaci o zdrojích (a tedy se značkováním vnějším), ale i struktuře, ba přímo povaze jazykových jednotek v něm (a tedy se značkováním vnitřním) (Čermák 2006: 10).

Językoznawstwo korpusowe (...) zajmuje się analizą języka zgromadzonego w korpusach językowych czyli komputerowych zbiorach autentycznych tekstów językowych, mówionych i pisanych, reprezentujących różne odmiany, style i typy tekstu (Lewandowska-Tomaszczyk 2005: 11).

Kiedy mówi się o „cyfrowych zbiorach tekstów”, korpusy przychodzą na myśl wręcz automatycznie. Przez „korpus” rozumiem: a) zbiór tekstów; b) zebranych przez kogoś; c) pod pewnym kątem; d) dla określonych celów (najczęściej badawczych); e) mający postać cyfrową – i w związku z tym; f) wyposażony w urządzenie przeszukujące. Wynika z tego, że nie każdy zbiór tekstów jest korpusem (Chlebda 2013: 9).

Zgodnie z powyższymi definicjami należy stwierdzić, iż korpusem językowym nazwiemy cyfrowy zbiór anotowanych tekstów naturalnych, reprezentujących zbiory pisane i transkrypcje języka mówionego, charakteryzujących się różnorodnością stylów, wariantów i typów funkcjonalnych. Korpus różni się od zwykłego zbioru tekstów obligatoryjnymi kryteriami ich wyodrębniania i uporządkowania, do których należą:

- a) **Ilość**. Jest wartością domyślną i polega na gromadzeniu ogromnych ilości danych w korpusie w postaci zbiorów tekstowych. Rozwój technologii implikuje coraz większą liczbę korpusów i rejestrowanych w nich form wyrazowych.
- b) **Jakość**. Zawarte w korpusach dane odzwierciedlają rzeczywisty proces komunikacji w określonym języku.
- c) **Prostota**. Wiąże się z możliwością kodowania znaków przez system komputerowy. W wyniku rozwoju techniki komputerowej powstają ulepszone systemy kodowania znaków, jak np. Unikod⁵.
- d) **Dokumentacja**. Oznacza staranne przechowywanie danych językowych w formie odrębnej bazy danych.
- e) **Reprezentatywność**. Celem korpusu jest odzwierciedlenie różnych form gatunkowych i funkcjonalnych języka w zrównoważonych proporcjach, a także posiadanie wystarczającej próby tekstów i ich autorów w celu przeprowadzenia rzetelnych badań określonych zjawisk językowych (Kozera 2018: 20).
- f) **Przechowywanie danych na nośniku elektronicznym**. Zaletą gromadzenia baz tekstowych w formacie cyfrowym jest zarówno możliwość szybkiego przeszukiwania danych, jak i natychmiastowego wzbogacenia zbioru o nowe informacje bądź uwagi.
- g) **Referencyjność**. Ciągła baza danych umożliwia szeroką dostępność dla innych badaczy, a także porównywanie i odwoływanie się do uprzednio osiągniętych wyników i analiz (McEnery i Wilson 2001: 32).

Niektórzy badacze wyodrębniają jeszcze jedno kryterium, a mianowicie *skończony rozmiar korpusu*, który wymagałby wstępnego zaplanowania jego finalnej objętości, a po jej

⁵ Komputerowy system kodowania zestawu znaków uwzględniającego literę większości pism świata.

osiągnięciu zamknięcie zbioru tekstowego. Stała wielkość korpusu umożliwia obiektywną obserwację analizowanego zagadnienia czy zbadanie faktów językowych w określonym wycinku czasowym. W opozycji funkcjonują jednak korpusy monitorujące, których struktura i rozmiar ulegają ciągłym zmianom, nienaruszającym jego reprezentatywności i proporcjonalności zawierających się w nich tekstów.

3.2 Typy korpusów

We współczesnej lingwistyce korpusowej istnieją dwa główne sposoby podziału korpusów na określone klasy:

- 1) Rozgraniczenie baz tekstowych, odnoszących się do języka ogólnego (*korpus ogólny*) od korpusów opisujących konkretny wycinek rzeczywistości językowej (np. gatunek, styl, język określonej grupy społecznej – *korpus specjalistyczny*),
- 2) Podział ze względu na zastosowany rodzaj anotacji (większość korpusów posiada anotację morfologiczną lub syntaktyczną). Wielość problemów badawczych i podejmowanych zagadnień językowych implikują obecność szerokiego repertuaru powstałych baz tekstowych, z których każda służy jako materiał dla osiągnięcia konkretnych celów. Poniższa tabela służy zaprezentowaniu najbardziej rozpowszechnionych korpusów i ich podziałów:

Cecha korpusu	Typ korpusów
Rodzaj danych	Pisane
	Mówione
	Mieszane
Język tekstów	Rosyjski, Angielski, Czeski, Polski...itp.
Paralelność	Jednojęzyczne
	Dwujęzyczne
	Wielojęzyczne
Literackość, Specyficzność	Języka literackiego
	Dialektów
	Języka potocznego
	Terminologiczne
	Mieszane
Gatunek	Literackie
	Folkowe

	Dramat
	Publicystyka
Dostępność	Ogólnodostępne
	Komercyjne
	Niedostępne
Przeznaczenie (Zastosowanie do...)	Badań
	Ilustrujące
Dynamiczność	Monitorujące
	Referencyjne
Anotacja	Anotowane
	Nieanotowane
Rodzaj anotacji	Morfologiczna
	Syntaktyczna
	Semantyczna
	Prozodyczna itp.
Objętość tekstów	Korpus, zawierający teksty w całości
	Próbkowany
Chronologia	Synchroniczne
	Diachroniczne

3.3 Rodzaje analizy korpusowej

W zależności od charakteru podejmowanego badania i założonych celów analizy wyróżniamy trzy podstawowe metody badawcze w lingwistyce korpusowej. Traktowanie danych korpusowych jako całości i zawieranie się we wnioskach wyłącznie poświadczeń korpusowych jest charakterystyczne dla analizy *Corpus-driven*, która wykorzystuje bazy tekstowe w celu skonstruowania zasad i ustalenia wyjątków w języku. Spora krytyka tej metody wynika z faktu, że zakłada ona świadome porzucenie założeń teoretycznych podczas analizy, co nie zawsze jest możliwe — chociażby ze względu na konieczność wcześniejszego określenia przedmiotu badań. Powtarzalność występujących wzorów oraz rozkład częstości w korpusach jest podstawą do wyłożenia twierdzeń w naszym rozwiązaniu metodologicznym. Podejście *Corpus-based*, wykorzystuje korpusy do wyłożenia, sprawdzenia lub egzemplifikacji teorii oraz opisów, które były sformułowane przed dostępem do wielkich korpusów (Tognini-Bonelli 2001: 65). Założenia teoretyczne są formułowane przed analizą i nie ulegają zmianie podczas prowadzenia badań. Niektórzy lingwiści tym samym pojęciem określają każde badanie z

wykorzystaniem korpusu bez względu na zastosowaną w nim metodę. Z kolei metoda *Corpus-assisted analysis* to jednoczesna analiza danych korpusowych oraz pozakorpusowych (np. wywiady, ankiety itp.).

3.4 Podstawowe narzędzia badawcze korpusów

3.4.1 Wyszukiwanie interaktywne

Głównym narzędziem obróbki danych korpusowych pozostaje funkcja wyszukiwania ciągów znaków, która umożliwia zarówno wpisanie zapytania w postaci wyrazu, frazy bądź wyrażenia regularnego, jak i późniejsze otrzymanie listy poświadczeń (*konkordancji*), zawierających wyszukiwaną przez użytkownika jednostkę. Podstawą skutecznej pracy z wyszukiwarką jest uprzednia znajomość zasad zapytań wybranego korpusu, na podstawie których powstał system znaczników, co umożliwi pobranie rzeczywistej listy poświadczeń wybranej jednostki. Poszczególne korpusy znacząco różnią się funkcjonalnościami, a niekiedy różnice są obserwowane w ramach tej samej bazy tekstowej, np. w dwóch wyszukiwarkach Narodowego Korpusu Języka Polskiego, tj. Poliqarp i PELCRA, tylko druga uwzględnia pisownię jednostki zarówno małą, jak i dużą literą (Hebal-Jezierska 2013: 24). W ostatnich latach udostępniono również nową wyszukiwarkę korpusową MTAS, umożliwiającą przeszukiwanie NKJP z uwzględnieniem rozszerzonych poziomów anotacji, takich jak zależności składniowe oraz jednostki nazewnicze. Istotne na tym etapie jest również zwrócenie uwagi na obszar bazy tekstowej, w której wpisywane jest zapytanie, gdyż niektóre korpusy umożliwiają wyszukiwanie nie tylko w zasobach ogólnych, ale i wybranych podkorpusach, np. NKJP – w korpusie zrównoważonym, NKJR – w szesnastu podkorpusach, a w Narodowym Korpusie Języka Czeskiego – w osiemnastu. Otrzymane wyniki przeważnie zaprezentowane są w postaci listy wszystkich zastosowań zadanej jednostki w kontekście. Najpopularniejszą formą jest format KWIC (*Key word in context*), w którym wyróżnione słowo znajduje się w prawo- i lewostronnym otoczeniu kontekstowym, pochodzącym z autentycznych tekstów.

3.4.2 Anotacja

Mnogość i różnorodność podejmowanych problemów badawczych wymaga nie tylko dostępu do dużych baz tekstowych, ale również dodatkowych metainformacji o zebranych w korpusach tekstach i wybranych jednostkach leksykalnych. Służy temu tzw. **anotacja** (ros. *разметка*, czes. *anotace*), przypisująca tekstom i ich elementom specjalne oznakowania (tagi). Dostarcza ona informacje o autorze, tytule, roku czy gatunku wybranego fragmentu tekstu

(*anotacja zewnętrzna*) bądź dotyczy cech leksykalnych, składniowych czy gramatycznych konkretnego leksemu (*anotacja wewnętrzna*) (Hebal-Jezińska 2014: 7). Wybranie metody anotacji jest kluczowym zagadnieniem przy opracowywaniu korpusu i musi być związane z jego docelowym przeznaczeniem i specyfiką zebranych w nim tekstów. Najczęściej stosowana metoda formalno-morfologiczna polega na automatycznym systemie anotacji, który przypisuje jednostce określone znaczniki, niezależnie od cech jej użycia → «Формально-морфологический подход позволяет разметить огромные массивы текстов без участия человека (программа-парсер приписывает информацию, руководствуясь электронными морфологическими словарями-указателями словоформ)» (Ляшевская, Плунгян и Сичинава 2005: 111-112). Inna metoda anotacji sprowadza się do określenia znaczenia semantycznego każdego wyrazu czy frazy i ich użycia w tekście, co jest zadaniem problematycznym w przypadku dużych zasobów tekstowych. Do najbardziej rozpowszechnionych anotacji wewnętrznych (językowych) należą:

- a) **Anotacja morfologiczna.** Jest uważana za podstawę dalszych form analizy i służy dla oznaczania części mowy i kategorii gramatycznych jednostek, np. **kot** [kot:subst:sg:nom:m2].
- b) **Anotacja syntaktyczna.** Dany typ znaczników opisuje relacje składniowe pomiędzy jednostkami i struktury układów składniowych, np. zdania podrzędnego, frazy czasownikowej itp. Jest wynikiem parsowania, przeprowadzonego na podstawie danych analizy morfologicznej, np. **Kamkoli** [Adv] *jsem* [AuxV] *přišel* [Obj].
- c) **Anotacja semantyczna.** Oznacza kategorie semantyczne, do których należy określony wyraz bądź fraza, oraz bardziej szczegółowe podkategorie dla określenia znaczenia jednostki, np. **солнце** [космические объекты, предметные имена, горизонтальные поверхности, пространство и место, сферические объекты].

Należy zauważyć, że istnieją również korpusy nieanotowane, których zaletą są autentyczne informacje autorskie zawarte w tekście bez informacji dodatkowych, które mogłyby się przyczynić do zniekształcenia odbioru tekstów poprzez zastosowanie różnych koncepcji znakowania. Anotacja zawsze zostaje powiązana z konkretnym tekstem i jest od niego zależna, jednak wnosi znaczącą wartość dodaną z uwagi na możliwość podjęcia bardziej szczegółowych badań językoznawczych (Čermák 2017: 33). Należy mieć również na uwadze wysokie prawdopodobieństwo wystąpienia błędów w anotacji, które najczęściej dotyczą form homonimicznych i synkretycznych – w takiej sytuacji prawidłowość oznaczeń należy sprawdzić za pomocą innych procedur.

3.4.3 Lematyzacja

Proces *lematyzacji* (*hasłowania*) polega na redukcji form fleksyjnych jednostki do jej formy podstawowej, tzw. *lemmy*. Zastosowanie danej procedury jest szczególnie istotne w przypadku języków fleksyjnych z uwagi na możliwość występowania jednostki w wielu formach i umożliwia eliminację potencjalnych nieprawidłowości w listach frekwencyjnych. Jednym z zagadnień najbardziej kłopotliwych dla automatycznej lematyzacji pozostają formy homonimiczne, z powodu których narzędzie przypisuje im nieprawidłowe właściwości gramatyczne i formę wyjściową, np.:

*Czasem na parkiecie ledwo zostawały dwie, trzy pary, a większość tam jadła, **pila*** [piła:subst:sg:nom:f].

*Bo słyszał o tym pierwszy raz. Dąb się zamyślił: "Rzecz zawiła, By drzewo ściąć - potrzebna **pila**, Piły wyrabia się ze stali, Stal daje huta, i tak dalej.* [piła:subst:sg:nom:f]

*Pan to pięknie akcentuje, a Fryderyk nie potrafi, więc niech ma Sonię. Wieki cale nie **piłam** tak cudownej herbaty.* [pić:praet:sg:f:imperf]

Lematyzacja w korpusie Poliqarp NKJP niewłaściwie wyprowadza formę podstawową dla wyrazu **pila**, będącym w pierwszym przykładzie formą trzeciej osoby liczby pojedynczej czasownika **pić**. Jak można zauważyć, błędna lematyzacja pociągnęła za sobą również mylną anotację jednostki i przypisała jej cechy gramatyczne rzeczownika **pila** (przykład drugi). Obserwacja innych konkordancji pozwala stwierdzić, iż korpus prawidłowo przypisał lemat **pić** wraz z oznaczeniami właściwymi dla czasowników w przypadku każdej odmienionej formy czasownika oprócz trzeciej osoby liczby pojedynczej, która stanowi homonimię dla ostrego narzędzia. Problem z błędnym hasłowaniem komplikuje się jeszcze bardziej w przypadku wyrażen, należących do tej samej części mowy, ale posiadających inne znaczenie zależne od kontekstu użycia. Rozwiązaniem może być uprzednie zastosowanie anotacji semantycznej, w przypadku której automatyczna obróbka danych umożliwi rozróżnienie znaczenia w kontekście szerszej wypowiedzi. Opisy procesów lematyzacji nierzadko powiązywane są bezpośrednio z procedurą wyodrębnienia przez korpus tematu fleksyjnego i słowotwórczego wyrazu, czyli techniką **stemmingu**. Algorytm stemmingu wykazuje wysoką przydatność w przypadku

wyrazów, których temat nie pokrywa się z morfologicznym rdzeniem słowa (Жердева i Артюшенко 2016: 131).

3.4.4 Kolokator

Niektóre z korpusów są wyposażone również w narzędzie *kolokatora*, które umożliwia odnalezienie w zasobach tekstowych najbardziej łączliwych słów dla wybranej jednostki, czyli *kolokacji*. Reprezentują one kombinacje słów, wykazujących duże prawdopodobieństwo współwystępowania obok siebie i uwzględniają cechy idiomatyczne, frazeologiczne połączenia wyrazowe oraz względną reprodukowalność w tekstach. Wzorce współwystępowania słów i ich powtarzalność są podstawą dla tworzenia słowników specjalistycznych, np. idiomatycznych bądź kolokacyjnych, a także w procesie nauczania języka obcego. W celu sprawdzenia łączliwości słów najczęściej wykorzystuje się wzory obliczania występowania, tj. MI (mutual information) i Testu Z⁶. Wyniki zastosowania tych metod są wykorzystywane do wyodrębniania jednostek wielowyrazowych i ich poszczególnych znaczeń, właściwych dla kontekstu, w którym występują. Kolokacje mogą być udostępniane przez korpus w formie listy połączeń wyrazowych dla zadanej jednostki, a wyekscerpowane przykłady odsyłają użytkownika do konkretnych poświadczeń tekstowych. Poniżej zaprezentowano pierwszych dziesięć kolokacji dla wyrazu *człowiek* w kolokatorze NKJP:

Ośrodek kolokacji występuje w korpusie 50395 razy.		Podajemy wyniki na podstawie współwystąpień od 0 do 10000.	Znaleziono 579 kolokacji spełniających zadane kryteria.	
#	Kolokacja	Pasujące współwystąpienia (kliknij na frekwencję, aby wyświetlić przykłady)	Ogółem	Chi ²
1.	młody	młody _____ człowiek (298), człowiek _____ młody (15).	313	43,755.4
2.	porządny	porządny _____ człowiek (35), człowiek _____ porządny (4).	39	7,494.79
3.	poszczególony	poszczególony _____ człowiek (8).	8	5,691.73
4.	współczesny	człowiek _____ współczesny (37), współczesny _____ człowiek (19).	56	5,296.71
5.	niemłody	człowiek _____ niemłody (4), niemłody _____ człowiek (2).	6	3,079.14
6.	światły	światły _____ człowiek (5), człowiek _____ światły (4).	9	2,094.09
7.	normalny	normalny _____ człowiek (39), człowiek _____ normalny (4).	43	1,998.56
8.	dorosnąć	dorosły _____ człowiek (18), człowiek _____ dorosły (9).	27	1,771.72
9.	dorość	dorosły _____ człowiek (18), człowiek _____ dorosły (9).	27	1,771.72
10.	dorosły	dorosły _____ człowiek (18), człowiek _____ dorosły (9).	27	1,771.72

Inny sposób zwizualizowania najczęstszych połączeń wyrazowych ma zastosowanie w Narodowym Korpusie Języka Czeskiego (NKJCz), w którym zapytanie *člověk* otwiera dostęp do mapy kolokacji, wykazujących się najwyższą frekwencją łączliwości z jednostką wyjściową:

⁶ Test Z – badanie uwzględniające średnią arytmetyczną i wartość odchylenia standardowego.



Source: [syn2015](#), more detailed information:
[Kolokace v SYN2015 \(KonText\)](#)¹⁴

3.4.5 Zastosowanie korpusów

Dostępność zawartych w korpusach dużej liczby tekstów wraz z przypisanymi im informacjami metatekstowymi sprawia, że zasadniczym przedmiotem badań pozostaje analiza wszelkiego rodzaju faktów językowych. Głównymi odbiorcami korpusów pozostają lingwiści, w przypadku których początkowe wykorzystanie baz tekstowych sprowadzało się do obliczania frekwencji występowania poszczególnych struktur zdaniowych. Wraz z rozwojem językoznawstwa korpusowego możliwość ekstrakcji danych z korpusów umożliwiła ich wykorzystanie w celach automatycznego przetwarzania tekstu i do opracowywania słowników, glosariuszy i tezaurusów. Korpusy stanowią również podstawę eksperymentalną dla weryfikacji założonych uprzednio hipotez i sprawdzenia teorii, powstałych jeszcze przed pojawieniem się elektronicznych zbiorów tekstów. Lingwistyka stosowana efektywnie użytkuje korpusy jedno- i wielojęzyczne dla celów dydaktycznych, głównie w procesie nauczania języka obcego. Ponadto korpusy umożliwiają również skuteczny rozwój i kształtowanie kompetencji poznawczych uczniów poprzez pracę z kolokacjami, zapewniającymi skuteczną obserwację jednostek leksykalnych w ich najczęstszym otoczeniu tekstowym, a przykładem wykorzystania korpusów w tym obszarze może być powstanie i rozwój korpusów uczniowskich. Są one zbiorem danych językowych od osób uczących się określonego języka jako języka obcego i znajdują swoje zastosowanie w teorii przyswajania języka drugiego oraz szeroko pojętej glottodydaktyce (Lewandowska-Tomaszczyk 2005: 203). W lingwistyce komputerowej korpusy służą do identyfikacji i wykorzystywania zawartych w nich tekstów do opracowania wzorców statystyczno-językowych, stosowanych w procesie komputerowego przetwarzania modeli języka, tłumaczenia maszynowego czy automatycznego rozpoznawania mowy. Zawarta w korpusach języka mówionego metainformacja, dotycząca szczegółowych informacji na temat nadawcy wypowiedzi (płeć, wiek, wykształcenie itp.)

umożliwia wykorzystanie baz tekstowych w badaniach z zakresu nauk społecznych. Przedstawione obszary stosowania korpusów powinny skłaniać do wniosku, że współcześnie mniejszość stanowią badania językoznawcze niewykorzystujące danych korpusowych.

3.5 Narodowy Korpus Języka Polskiego

3.5.1 Ogólna charakterystyka NKJP

Pierwszy zróżnicowany i reprezentatywny korpus polszczyzny zaczął powstawać z inicjatywy Prezydium Komitetu Językoznawstwa Polskiej Akademii Nauk w 2006 roku, dając tym samym początek NKJP, nad którym prace prowadziły cztery ośrodki badawcze: Instytut Podstaw Informatyki PAN, Instytut Języka Polskiego PAN, Wydawnictwo Naukowe PWN i Zakład Językoznawstwa Komputerowego i Korpusowego Uniwersytetu Łódzkiego (Bańko i Górski 2014: 11). Każdy z zespołów dysponował zgromadzonymi wcześniej bazami tekstowymi, które dzięki połączeniu i uzupełnieniu pozwoliły stworzyć anotowany morfosyntaktycznie korpus o objętości 1,8 miliarda segmentów. W trakcie prac projektowych opracowano również liczne narzędzia do oznakowania i przetwarzania tekstów, takie jak tager PANTERA, system ujednoznaczniania sensów słów, kolokator czy wyszukiwarka danych mówionych. W 2011 roku korpus stał się największym anotowanym zbiorem danych języka polskiego, a zawarte w nim materiały stały się podstawą dla opracowania Wielkiego Słownika Języka Polskiego IJP PAN czy skonstruowania Poradni Językowej PWN, niejednokrotnie odsyłająca użytkownika do konkretnych danych korpusowych NKJP (Lewandowska-Tomaszczyk i in. 2012: 9).

Strukturę korpusu stanowią dwie główne bazy tekstowe, tj. Korpus główny (1,8 mld segmentów), Korpus zrównoważony (300 mln segmentów), a także wyszukiwarki Poliqarp (mieści w sobie mniejsze podkorpusy), PELCRA (umożliwiająca wybór konkretnego zbioru) oraz MTAS. Najwcześniejsze teksty korpusu są datowane na przełom XVIII i XIX wieków, jednak przeważająca część zbiorów zawiera teksty prasowe, internetowe czy sprawozdania z posiedzeń parlamentu. Typologia zawierających się w NKJP tekstów została tak skonstruowana, aby każdy tekst mógł zostać przypisany do odpowiedniej kategorii i tym samym zostać właściwie zdefiniowany. Do wspomnianych typów należą: 1) literatura piękna (proza, poezja, dramat), 2) literatura faktu, 3) publicystyka i wiadomości prasowe, 4) typ naukowo-dydaktyczny, 5) typ informacyjno-poradnikowy, 6) książka niebeletrystyczna nieskalsyfikowana, 7) inne teksty pisane (np. ogłoszenia, reklamy, teksty instruktażowe), 8) listy, 9) Internet (np. chaty, strony WWW), 10) teksty mówione konwersacyjne, 11) teksty

mówione medialne, 12) teksty quasi-mówione (Bańko i Górski 2014: 13). Większość badań i kwerend jest przeprowadzana na korpusie zrównoważonym, co naturalnie jest związane z możliwością otrzymania najbardziej wiarygodnych wyników z uwagi na proporcjonalny udział tekstów. Procentowy udział każdego typu tekstu w korpusie przedstawiono poniżej:

Rodzaj tekstów	Udział w NKJP
Książki	29,0%
z tego:	
<i>Literatura piękna</i>	16,0%
<i>Literatura faktu</i>	5,5%
<i>Książki naukowo-dydaktyczne</i>	2,0%
<i>Książki i prasa informacyjno-poradnikowe</i>	5,5%
Prasa	50,0%
z tego:	
<i>Gazety</i>	26,0%
<i>Periodyki</i>	24,0%
Inne teksty pisane (urzędowe, listy)	4,0%
Internet (blogi, fora, strony www)	7,0%
Teksty mówione (konwersacyjne naturalne i medialne, protokoły sejmowe)	10,0%

Dążenie do zrównoważenia stylowo-gatunkowego, chronologicznego i tematycznego pozwoliło stworzyć podstawy dla dalszego rozwoju Korpusu Języka Polskiego, którego szczególnym osiągnięciem była rejestracja tekstów mówionych. Pozyskiwanie, transkrypcja i anotacja języka mówionego jest procesem bardziej pracochłonnym i kosztownym niż opracowywanie tekstów pisanych. Niemniej jednak uwzględnienie tej odmiany języka jest uzasadnione jej największym udziałem pod względem ilości konstruowania wypowiedzi. Włączenie do korpusu tekstów mówionych umożliwiło ich analizę leksykalno-składniową czy pragmatyczną, co stało się szczególnie istotne z uwagi na fakt pomijania danych mówionych w badaniach lingwistycznych czy opracowaniach leksykograficznych. Pomimo dominacji w tekstach mówionych zapisów z posiedzeń Sejmu nie można odmówić ich użyteczności w analizie współczesnego języka potocznego.

3.5.2 Wyszukiwanie w NKJP

Jak już zostało wspomniane, Narodowy Korpus Języka Polskiego został wyposażony w trzy wyszukiwarki:

- 1) PELCRA - charakteryzuje się prostą składnią zapytań i pozwala analizować pozyskane kwerendy głównie pod kątem leksykalno-semantycznym,
- 2) Poliqarp – skupia się na gramatycznym aspekcie badań i wykorzystuje pojęcie segmentu, a nie wyrazu (np. *przyszł em*),

3) MTAS - oparta na języku zapytań CQL umożliwia analizę danych na podstawie segmentów i tagów morfosyntaktycznych. Stanowi rozwinięcie funkcjonalności Poliqarp.

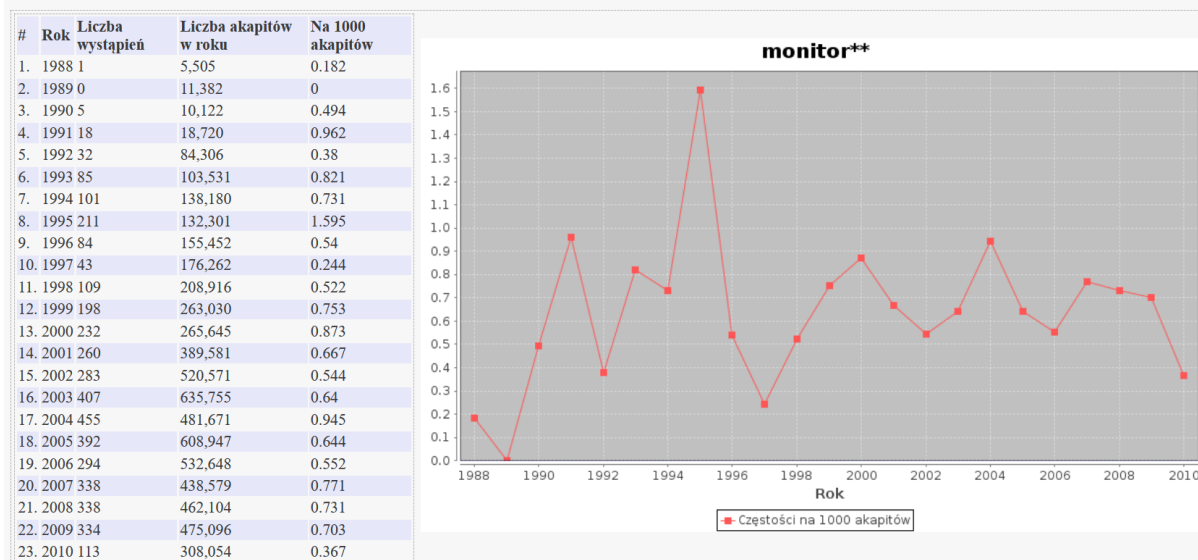
Z racji podejmowanych w tekście badań leksykologicznych i złożonego systemu języka zapytań Poliqarp i MTAS, przeznaczonych w głównej mierze do morfologiczno-gramatycznej analizy jednostek, główna uwaga skupi się na wyszukiwarce PELCRA. Umożliwia ona wpisanie zapytania o dokładną formę zadanej jednostki, jej odmiany bądź następującego po wyrazie ciągu słów przy użyciu symboli zastępczych, tj. gwiazdki i pytajnika. Przy zastosowaniu przykładowego zapytania *podręcznik*** wynikami kwerendy będą konkordancje, uwzględniające wszystkie formy odmiany tego leksemu. Należy zaznaczyć, iż system nie tylko nie rozróżnia małej i dużej litery zapytania, lecz także ilustruje w kontekstach formy homonimiczne, które mogą istotnie przyczynić się do otrzymania wyników nadmiarowych i niewiarygodnych. Wartościowym dla badań ciągów wielowyrazowych czy frazeologizmów jest możliwość ustalenia określonej odległości pomiędzy jednostkami (opcja *Maks. odległość*) bez uwzględnienia wartości *Zachowaj szyk*, której wyłączenie daje prawdopodobieństwo uzyskania większej liczby wyników. Dodatkowe zastosowanie potrójnego znaku podkreślenia w zapytaniu przyczynia się do rozdzielenia wariantywnych składników sekwencji, np. kwerenda *wierzyć**__chcieć*** przy zmiennym ustalaniu wartości odległości i zachowania szyku przyniosła następujące formy ciągu: *Nie chce się wierzyć; Chcesz, to wierz; Chciałbym w to wierzyć; Wierzcie sobie, w co chcecie* itp. Poniżej przedstawiono rozszerzony formularz wyszukiwarki wraz z dodatkowymi funkcjami:

Maks. odstęp: 0		Zachowaj szyk: ☒	Wyniki: 100	Czas	Profil	Excel	URL
Ukryj opcje							
Sortowanie:	1: środek		2: środek				
Grupowanie:	---		:	1			
Typ (?) :	--- typ_fakt typ_inf-por typ_konwers						
Kanał (?) :	--- kanał_internet kanał_książka kanał_mowiony						
Data:	od RRRR		do RRRR				
Podkorpus (?) :	Zrównoważony						
Tytuł źródła (np. książki, gazety):							
Pomiń źródła (np. książki, gazety):							
Tytuł tekstu:							
Wymagane wyrazy kontekstowe:							
Niedopuszczalne wyrazy kontekstowe:							
SZUKAJ							

Opcja *Sortowanie* umożliwia wyświetlanie listy poświadczeń według słowa kluczowego, lewej bądź prawej strony kontekstu, a także według źródła, daty, typu i kanału przekazu (np. prasa, książka, Internet lub mowa). *Grupowanie wyników* pozwala ograniczyć liczbę konkordancji z konkretnego źródła lub kanału w celu uniknięcia nadmiaru powtarzających się danych. Rozszerzone funkcje wyszukiwarki zawierają również możliwość selekcji tekstów przy użyciu klasyfikacji stylowo-gatunkowej, typów tekstów, kanałów czy daty i źródła tekstu. Istotnym w procesie przeprowadzania kwerend jest konieczność zwiększenia liczby wyszukiwanych wyników. Wyszukiwarka PELCRA przyjmuje automatyczną wartość 100 poświadczeń na stronie, co może prowadzić do wyświetlenia tylko części poświadczeń zadanej jednostki.

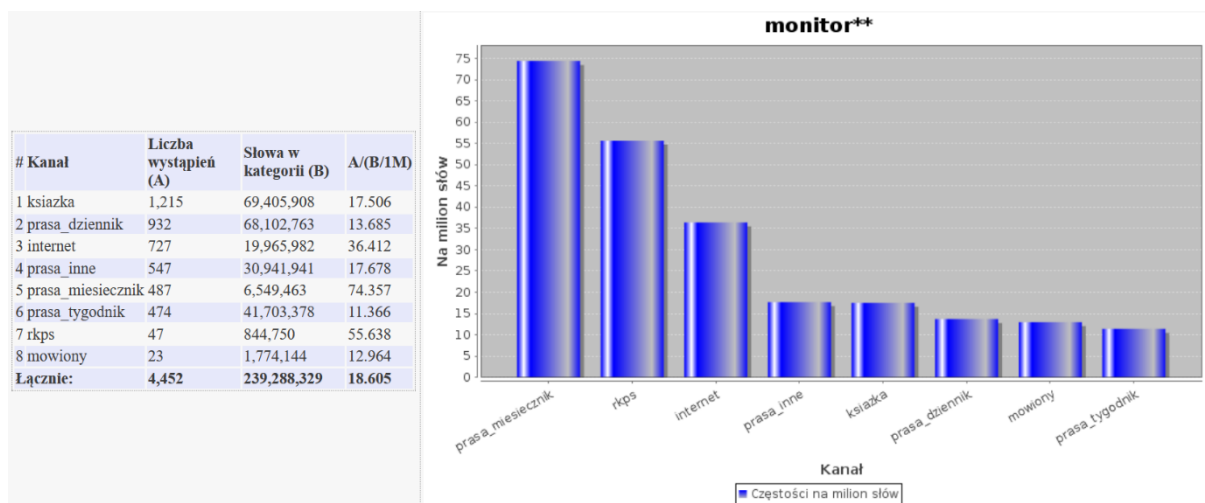
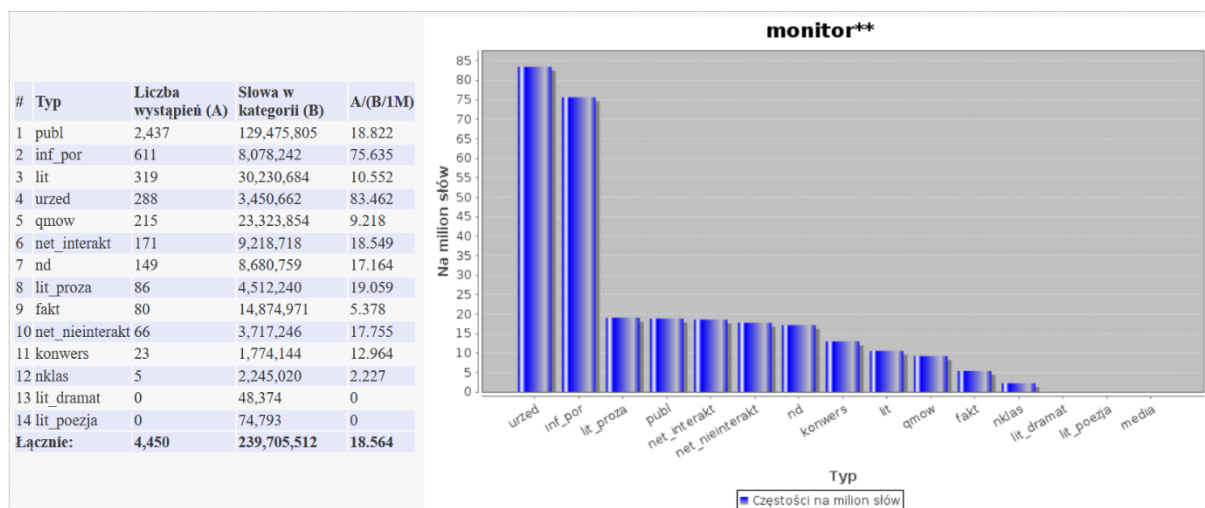
Dodatkową funkcją wyszukiwarki jest opcja *Czas*, podająca przybliżoną częstość użycia jednostki w okresie 1988–2010 i przedstawia wyniki w postaci tabeli i wykresu. Narzędzie umożliwia obserwację użycia jednostki czy związków wyrazowych w określonych przedziałach czasowych i analizę stopnia jej frekwencji w konkretnym typie tekstu i dyskursie. Należy pamiętać, iż frekwencje występowania są przeliczane na tysiąc akapitów, dlatego podczas analizy nie można się sugerować podaną liczbą wystąpień, tylko końcową wartością występowania w akapitach. Poniżej przedstawiono wyniki kwerendy *monitor*** z zastosowaniem wspomnianej funkcji:

Profil diachroniczny dla zapytania: *monitor***. Frekwencje obliczono w 1.477s.



Z kolei informację o dystrybucji stylistycznej słowa dostarcza opcja *Profil*, która z jednej strony wyświetla frekwencję jednostki w określonych typach tekstów i uwzględnia ich kryterium stylistyczne (np. publicystyka, teksty konwersacyjne), a z drugiej udostępnia podobne wyniki w podziale na kanały (np. książka, rękopis, prasa). Częstość występowania jest

przeliczana na milion słów, a przed rozpoczęciem kwerendy warto zapoznać się z objaśnieniem zastosowanych skrótów (np. *nd*, *nklas*, *qmow*) w zakładce *Pomoc*. Wyniki są podawane w formie tabel i wykresów o analogicznej konstrukcji jak w przypadku opcji *Czas*. Poniżej zaprezentowano profil występowania dla zapytania *monitor***.



3.5.3 Kolokator w NKJP

Wbudowane w wyszukiwarkę PELCRA narzędzie do automatycznego generowania listy typowych połączeń wyrazowych dla wybranego ośrodka kolokacji nosi nazwę *kolokator* i umożliwia ekskserpcję informacji o łączliwości poszczególnych słów. Użytkownik może ograniczyć wyszukiwanie kolokatów do konkretnych części mowy, określić maksymalne odstęp międzywyrazowe czy liczbę sąsiadujących z jednostką zadaną wyrazów. Ze względu na złożoność procesu ekstrakcji kolokacji wyszukiwarka może przeanalizować do 50 000 poświadczeń zadanego ośrodka kolokacji. Przykładowa tabela wyników dla zapytania *żartować*** została przedstawiona w poniższym zrzucie ekranu:

#	Kolokacja	Pasujące współwystąpienia (kliknij na frekwencję, aby wyświetlić przykłady)	Ogółem	Chi ²
1.	-	- żartuje (394), - żartował (260), - żartujesz (103), żartuje - (70), - żartują (48), żartujesz - (42), - żartowała (35), - żartowali (32), żartuję - (25), żartował - (25), - żartuję (20), - żartowałem (20), żartuj - (20), żartować - (16), żartują - (14), żartowałem - (14), - żartowałam (11), żartowałam - (6), - żartowano (5), - żartowaliśmy (5), - żartując (5), żartowała - (3), - żartuj (3), - żartujecie (3), żartowali - (3), żartowano - (3), - żartowaliśmy (2), żartując - (2), żartujecie - (2), - żartujemy (2), żartujecie - (1), żartowanie - (1), żartowaliśmy - (1), żartowania - (1), żartujmy - (1), - żartowanie (1), - żartowały (1), żartujemy - (1), żartowałeś - (1),	1202	18,300.35
2.	raczyć	raczy żartować (19), raczysz żartować (5), raczył żartować (2), żartować raczycie (1),	27	16,405.96
3.	ponuro	żartują ponuro (2), żartuje ponuro (1), ponuro żartowali (1), żartowano ponuro (1), żartował ponuro (1), ponuro żartują (1), ponuro żartuje (1),	8	8,060.77
4.	że	żartują że (101), żartował że (100), żartuje że (89), żartowali że (54), żartowano że (48), że żartuje (38), żartować że (23), żartuję że (23), żartowała że (19), żartujemy że (18), żartowaliśmy że (17), żartując że (15), żartowałem że (15), że żartuję (13), że żartował (8), żartowały że (7), że żartujesz (7), że żartują (5), żartowałam że (3), że żartowałem (3), żartuj że (3), że żartujemy (2), żartowaliśmy że (2), że żartowała (2), żartujesz że (1), że żartowanie (1), żartujecie że (1), że żartowałeś (1), że żartowali (1), żartowałaś że (1), że żartować (1), że żartowałam (1),	623	7,760.35
5.	chyba	chyba żartujesz (80), chyba żartuje (34), żartujesz chyba (24), chyba żartujecie (5), chyba żartował (2), żartuję chyba (2), chyba żartuję (2), żartuje chyba (2), chyba żartują (2), żartujecie chyba (1), żartowałeś chyba (1), żartował chyba (1), chyba żartowałeś (1),	157	7,489.28
6.	gorzko	gorzko żartował (4), żartuje gorzko (2), gorzko żartowali (1), gorzko żartuje (1),	8	7,255.11

Kolumna *Kolokacja* zawiera znormalizowane formy podstawowe odnalezionych kolokatów, a sąsiadująca z nią kolumna *Pasujące współwystąpienia* zawiera kombinacje kolokacyjne dla wszystkich odmian zadanej jednostki wraz z uwzględnieniem ich liczebności i odsyłaczem do listy konkordancji. Ogólna liczba występowania form w zasobach korpusowych podawana jest w kolumnie *Ogółem*, a szczegółowa frekwencja zostaje posortowana według wartości *Chi²*, która służy dla określenia istotności statystycznej kolokacji. Powyższy ekran wyników ogranicza się do zaprezentowania niektórych kolokatów, aczkolwiek analiza wszystkich współwystąpień pozwala wyodrębnić warianty danego emotywu: *Chyba żartujesz; Żartuję przecież; Dobra, żartowałem; Raczysz żartować?* itp. Kolokator umożliwia również odnajdywanie wielowyrazowych ośrodków kolokacji, szczególnie istotnych dla badań frazeologicznych i idiomatycznych. Przy ustawieniu kontekstu kolokacyjnego na jeden wyraz z prawej strony dopasowania i zadaniu zapytania *iść** do* narzędzie zwraca następujące przykłady kolokacji: *idź do diabła!; Idź do cholery!; Idź do piekła!* itp.

3.5.4 Anotacja w NKJP

NKJP nie ustępuje innym bazom tekstowym pod względem rozszerzonego systemu znakowania, który powstał w wyniku ręcznej anotacji morfoskładniowej. Proces ten polegał na podziale tekstu na zdania, segmenty, ich lematyzacji, przypisaniu części mowy i kategorii gramatycznych w korpusie jednomilionowym. Należy zauważyć, iż w wyszukiwarce Poliqarp nie istnieje pojęcie wyrazu, tylko segmentu oznakowanego jego hasłem podstawowym z przypisaniem części mowy i właściwości gramatycznych, np. oznakowanie wielowyrazowej jednostki *tańczyć do rana* przyjmuje następującą formę graficzną: **tańczyć** [tańczyć:inf:imperf] **do** [do:prep:gen] **rana** [rano:subst:sg:gen:n]. Korpus nie rozróżnia czasu przyszłego, zgodnie z czym formy *robię* i *zrobię* zaliczane zostają do czasu nieprzeszłego.

O szerokim repertuarze dostępnych znaczników świadczą również kategorie wokaliczności czy poprzyimkowości, a dzięki odróżnianiu rodzaju męskoosobowego od męskiego żywotnego możliwym staje się wyszukanie form z nieosobowym wykonawcą czynności. W odróżnieniu od innych korpusów narodowych w NKJP nazwa *anotacji wewnętrznej* ustąpiła terminowi *anotacja morfosyntaktyczna*, jednocześnie udostępniająca informacje o wszystkich formach zadanej jednostki i jej opisie gramatycznym. Dla procesu znakowania zostało wykorzystane pojęcie fleksemu, oznaczającego zbiór form jednolicie lub niemal jednolicie zróżnicowanych ze względu na właściwe im kategorie gramatyczne (Szałkiewicz, Przepiórkowski 2012: 62), np. klasa czasownik uwzględnia rozkazniki, bezokoliczniki, pseudoimiesłowcy, formy nieprzeszłe itp. Dalsze szczegółowe informacje dotyczące procesów lematyzacji, segmentacji czy anotacji tekstów pisanych podaje praca zbiorowa pt. *Narodowy Korpus Języka Polskiego* (red. A. Przepiórkowski, M. Bańko, R.L. Górski, B. Lewandowska-Tomaszczyk 2012), której lektura wydaje się niezbędna do efektywnego przeprowadzenia badań i analiz korpusowych.

Na oddzielne omówienie zasługuje system anotacji zastosowany w korpusie języka mówionego, w którym każda transkrypcja rozpatrywana jest jako jednolity tekst, a poszczególne wypowiedzi mówców są uznawane za podstawową jednostkę strukturalną. Uwzględniają one informacje dotyczące płci, wieku czy wykształcenia osoby mówiącej, a także podają dokładne źródło i datę nagrania. Wysoką użyteczność dla badań jednostek pragmatycznych w korpusie mówionym wykazują także dodatkowe metadane w postaci elementów, sygnalizujących zmianę tonu głosu czy towarzyszących wypowiedzi okolicznościom rozmowy. W określonych przypadkach transkrypcja zawiera również didaskalia, które podkreślają nagłą zmianę tematu rozmowy (<incident>) czy wskazują na

niezrozumienie przez osobę transkrybującą wyrazu bądź wiązki słów (<unclear>). Anotowany w korpusie dyskurs konwersacyjny odzwierciedla trudność wskazania granic zdań składniowych, zgodnie z czym nie należy traktować znaków interpunkcyjnych (kropek, pytańników) jako zakończenia myśli mówiącego. System znakowania wskazuje również na różnorodność podejmowanych tematów konwersacyjnych, co umożliwia dalszą obserwację czy wyodrębnienie zbitek leksykalnych, reproductów i potocyzmów współczesnej polszczyzny. Poniżej przedstawiono kontekst wystąpienia i anotację dla jednostki *odczep się* w korpusie mówionym NKJP:

Id		Tytuł	Źródło		Data nagrania	
190		Rozmowy młodzieży	PELCRA		2009	
Kontekst wystąpienia						
#	Id	Wypowiedź	Płeć	Wiek	Wykształcenie	
1.	31277	no kurde Basia nie chcę nie mogę no ja to już w ogóle nie mogę ja jadłam już	f	24	wy	
2.	31278	Ewa a ty zjesz jeszcze	f	23	sr	
3.	31279	mogę rybę zjeść	f	24	sr	
4.	31280	kapustkę . o Jezus . Sosna może piwa się napijesz	f	23	sr	
5.	31281	nie denerwuj mnie	f	24	wy	
6.	31282	aa bo ty nie /podchodzi do ciasta/ możesz a to jest jakieś bralni	f	23	sr	
7.	31283	nie . to jest . . odczep się od wysokości mojego /śmiech/ /śmiech/ ciasta to jest murzynek . ambasador taki	f	24	wy	
8.	31284	ambasador	f	23	sr	
9.	31285	no . toś się najadła . następnym razem ci przyniosę kanapkę	f	24	wy	

3.6. Korpus MoncoPL

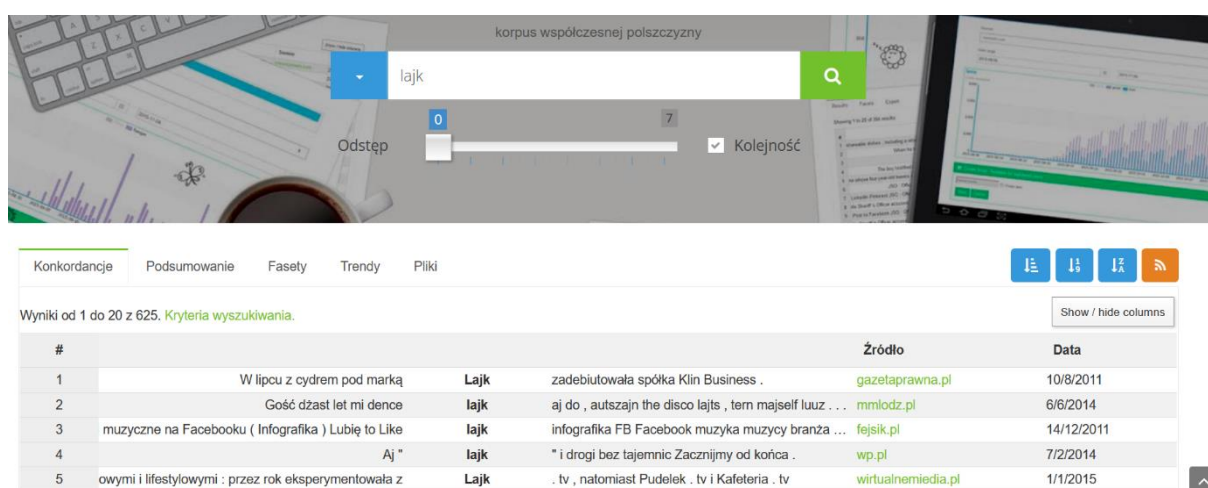
3.6.1 Ogólna charakterystyka

Elektroniczny zbiór danych tekstowych MoncoPL jest ogólnodostępnym korpusem monitorującym współczesnej polszczyzny, działającym od 2019 roku i udostępnionym w serwisie monco.frazeo.pl. Za źródło danych korpusowych posłużyły listy nagłówków artykułów, udostępnianych przez kanały RSS przez ponad 1500 serwisów informacyjnych, charakteryzujących się różnorodnością profili tematycznych. Autentyczne próbki tytułów i wpisów internetowych zostały uzupełnione o jednorazowe włączenie niektórych otwartych korpusów, z których największym jest Polski Korpus Sejmowy (Pęzik 2020: 134). Indeks wyszukiwarki liczy obecnie ponad 8 miliardów segmentów słów, ok. 450 tysięcy zdań, a niewątpliwą zaletą jego codziennej aktualizacji jest możliwość odnalezienia nie tylko wyrażen rzadkich, ale również neosemantyzmów o efemerycznym charakterze, np. *turlaj dropsa; eluwina; śpiulkolot*. Do pięciu największych źródeł tekstowych należą serwisy internetowe

gazeta.pl, tvn24.pl, interia.pl, onet.pl, gazetaprawna.pl, a średni przyrost danych wynosi ok. 87 tys. zdań dziennie. Ekstrakcję danych tekstowych automatycznie przeprowadza generyczny algorytm, uwzględniający również treść przypisanych artykułowi sekcji komentarzy, w których najsilniej przejawiają się współczesne tendencje językowe użytkowników polszczyzny, np. silne nasycenie wypowiedzi aktualnym dyskursem internetowym. Możliwość badania trendów i cykli frekwencyjnych w badaniach językowych pozwala stwierdzić, iż korpus MoncoPL stanowi w określonym stopniu uzupełnienie NKJP. Świadczy o tym nie tylko rozpoczęcie jego działalności w roku zamknięcia korpusu języka polskiego, ale również wyposażenie w narzędzie przeszukiwawcze, którego budowa opiera się na składni zapytań, właściwej dla wyszukiwarki PELCRA (Pęzik 2020: 135).

3.6.2 Wyszukiwanie w MoncoPL

Wyszukiwarka w korpusie MoncoPL cechuje się prostym interfejsem użytkownika, a funkcje narzędzia można podzielić na wyszukiwawcze i agregacyjne. Podstawowym zadaniem pierwszej z nich jest wyszukiwanie dokładnych słowoform bądź fraz, a także generowanie listy konkordancji w formie KWIC. Wyszukiwarka umożliwia również ustalenie liczby odstępów pomiędzy wyrazami w jednostkach wielowyrazowych i zachowanie ich określonej kolejności. Poświadczenia są uzupełnione o informacje z nazwą źródła, z którego pochodzi cytat wraz z datą publikacji tekstu, a także bezpośrednim odnośnikiem do pełnej wersji tekstu na oryginalnej stronie portalu. Poniżej ukazano ekran wyników wyszukiwania wyrazu *lajk* z kilkoma konkordancjami:

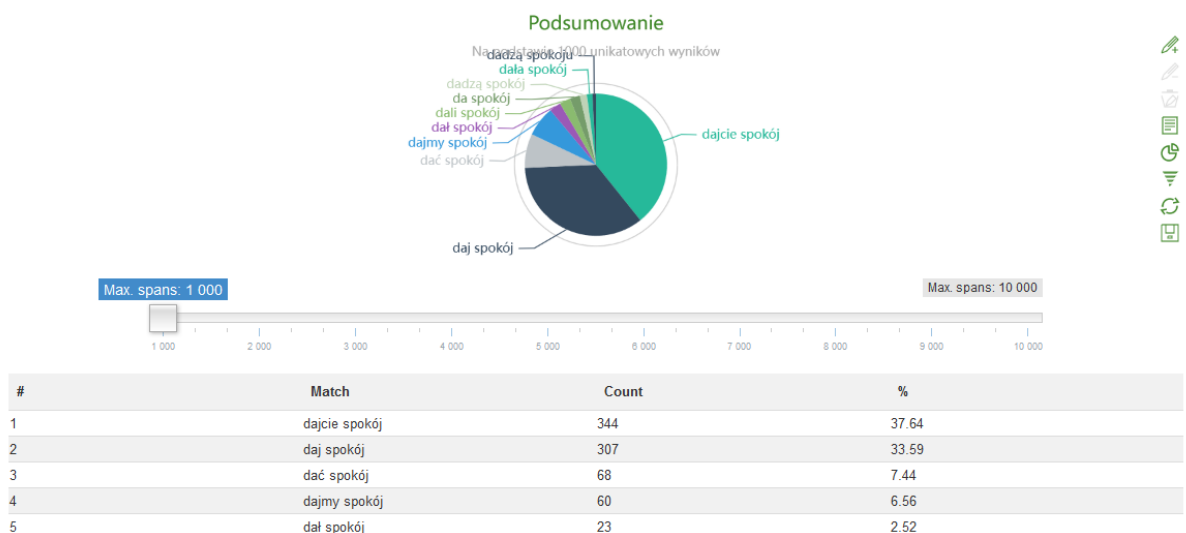


#	Wzrost	Lajk	Zadebiutowała spółka Klin Business	Źródło	Data
1	W lipcu z cydrem pod marką	Lajk	zadebiutowała spółka Klin Business .	gazetaprawna.pl	10/8/2011
2	Gość dziać let mi dence	lajk	aj do , autsajin the disco lajts , tern majself luuz . . .	mmlodz.pl	6/6/2014
3	muzyczne na Facebooku (Infografika) Lubię to Like	lajk	infografika FB Facebook muzyka muzycy branża . . .	fejsik.pl	14/12/2011
4	Aj *	lajk	" i drogi bez tajemnic Zacznijmy od końca .	wp.pl	7/2/2014
5	owymi i lifestylowymi : przez rok eksperymentowała z	Lajk	. tv , natomiast Pudelek . tv i Kafeteria . tv	wirtualnemedia.pl	1/1/2015

MoncoPL oferuje podobne możliwości w zakresie składni zapytań jak w przypadku NKJP poprzez wyszukiwanie wariantów fleksyjnych przy zadaniu lematu przy pomocy

konkretnych symboli, np. *łapa*** → *łapą, łapami, o łapie*; *duży** problem*** → *duże problemy, dużym problemom* itp. Istotnym instrumentem korpusu jest jego wyposażenie w operator wariantu (|), który jest wysoce przydatny w odnajdywaniu wariantów emotywów, np. aby odnaleźć jednocześnie wystąpienia fraz *dać spokój* lub *dać siana* można zastosować zapytanie w postaci *dać** spokój**|siano*** i przejść do zakładki Podsumowanie:

Unikatowych kontekstów pasujących do zapytania: 914. Maks. liczba kontekstów: 1000. Wszystkich kontekstów: 9302.



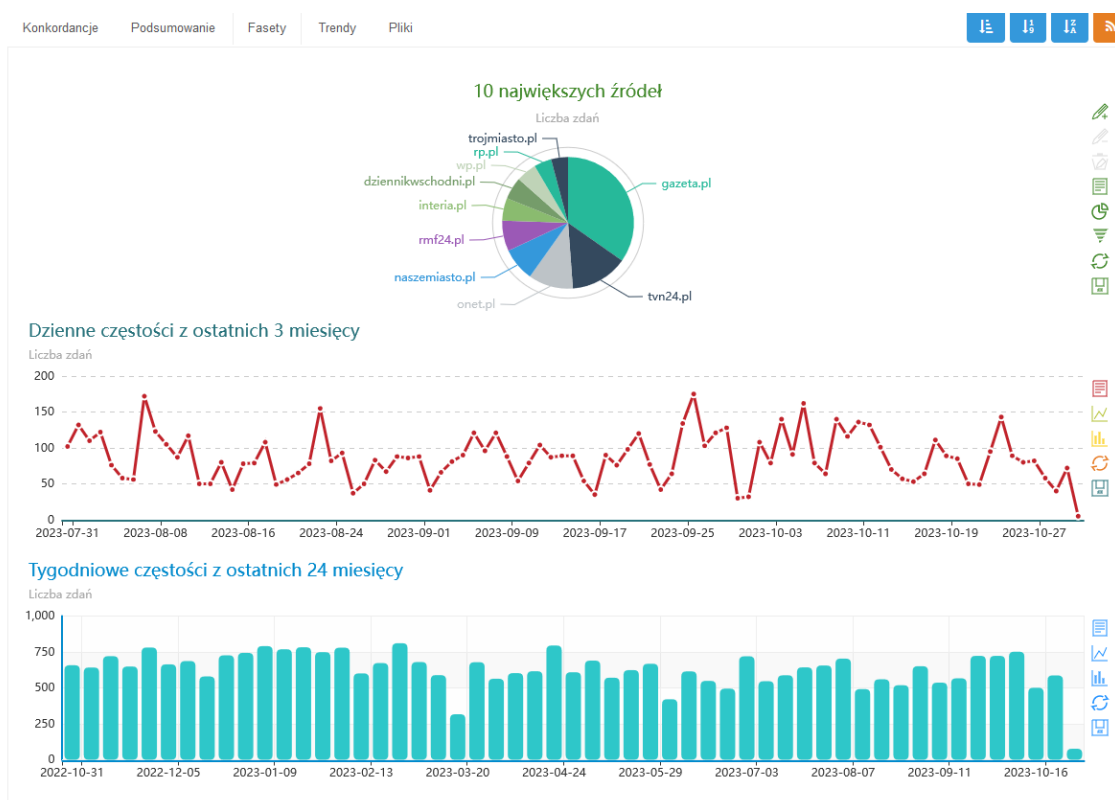
Z powyższej listy frekwencyjnej wynika, iż na podstawie próby o wielkości maksymalnej 1 tys. wystąpień najczęściej występującą formą jest wyrażenie *dajcie spokój*. Zaledwie pojedyncze poświadczenie (nieujęte w widocznej powyżej konkordancji) zarejestrowano dla formy *daj siana*, co najprawdopodobniej wynika z dużo niższej częstości jej użycia w porównaniu z konstrukcją *daj se siana*. Składnia MoncoPL umożliwia również definiowanie kategorii morfosyntaktycznych przy użyciu zapytania *<tag=XX>*, gdzie XX służy określeniu części mowy i innych kategorii gramatycznych, zaczerpniętych z tagsetu NKJP.

Założmy, że chcemy wyszukać przymiotniki występujące przed formami rzeczownika *kościół*. W tym celu należy skonstruować zapytanie *<tag=ADJ.*> kościół***, które oferuje następujące przykładowe wyniki: *chrześcijańskim kościele; barokowego kościoła; toruńskich kościołów* itp. Na osobną uwagę zasługuje wysoka przydatność funkcji *Odstęp* w procesie poszukiwawczym wariantów jednostek wielowyrazowych, która stwarza możliwość zaobserwowania w konkordancjach komponentów fakultatywnych konkretnej jednostki. Za przykład może posłużyć emotyw *Mi pasuje*, któremu można przypisać następującą definicję: *pot.* ‘używane w sytuacji, gdy informacja spełnia oczekiwania mówiącego; wyrażenie aprobaty

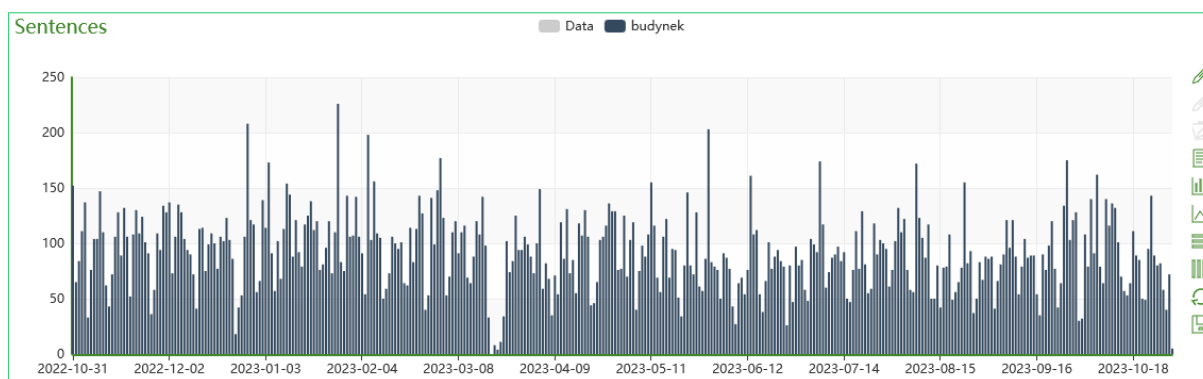
i zadowolenia’. Wpisanie tej jednostki do wyszukiwarki i ustalanie dla niej odstępów od 1–do 7 pozwoliło na wyodrębnienie z poświadczeń kolejnych wariantów: *mi całkiem pasuje*; *mi nie pasuje*; *mi to pasuje*; *mi bardzo pasuje*; *mi to nie pasuje*; *mi wyjątkowo nie pasuje*. Pozostałe konkordancje, w których zadana była liczba odstępów większa niż 2, rejestrowały głównie użycie jednostki w znaczeniu oceniającym obiekt bądź osobę, tj. *mi coś w <czymś/kimś> nie pasuje*.

3.6.3 Fasety i trendy

Funkcja *Fasety* udostępnia informacje o źródłach, w których zarejestrowana zadana jednostka jest najczęściej rejestrowana, a także umożliwia obserwację dziennych i tygodniowych częstości występowania wyrazu w określonych przedziałach czasowych. Dla zapytania *budynek* narzędzie wskazało dziesięć największych źródeł, umieszczonych na wykresie kołowym wraz z dodatkowymi informacjami, możliwymi do pobrania:

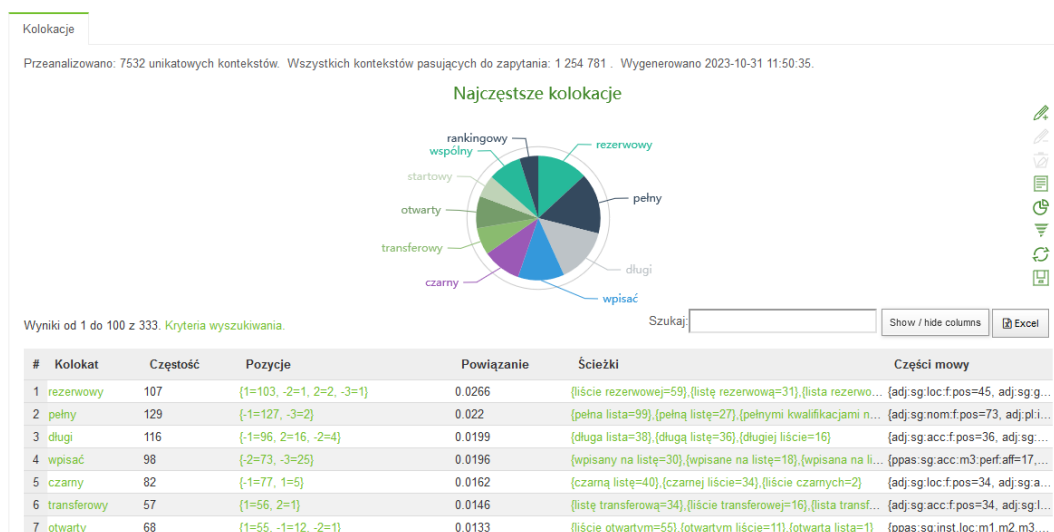


Narzędzie *Trendy* umożliwia z kolei badanie cykli frekwencyjnych, w których występowała konkretna jednostka, a także określenie źródła wraz z zakresem dat, pozwalającym na zanalizowanie frekwencji wyrazu w wybranym zbiorze i przedziale czasowym:



3.6.4 Kolokacje

Moduł ekstrakcji kolokacji znajduje się w zakładce *Narzędzia -> Kolokacje* i umożliwia odnalezienie frazemy czy idiomów, szeroko wykorzystywanych w opracowaniach z zakresu leksykografii jedno- czy dwujęzycznej. Podstawowe zapytanie stanowi tzw. ośrodek kolokacji, w przypadku którego można użyć dodatkowych filtrów morfosyntaktycznych lub określić maksymalną liczbę ekscerpowanych zdań dzięki parametrowi *Próbka*. Dodatkową zaletą narzędzia jest automatyczne usuwanie występujących w poświadczeniach duplikatów, które mogłyby doprowadzić do zaburzenia reprezentatywności otrzymanych wyników, a także możliwość określenia części mowy czy odległości kolokatów od ośrodka kolokacji. Poniżej przedstawiono przykładowe zapytanie na odnalezienie najbardziej łączliwych przymiotników dla jednostki *lista*:



Wśród innych dostępnych funkcji użytkownik może określić nazwy serwisów, z których mają być ekstrahowane kolokacje, zakres dat i minimalną częstość występowania. Na ekranie wyników widnieją również następujące metainformacje: 1) *Pozycje* – informacja o częstości występowania pozycji, na których wystąpił kolokat, np. *lista rezerwowa* – 103 wyniki dla drugiej pozycji, 2) *Powiązanie* – prosta miara powiązania oparta na współczynniku Dice'a, wyróżniająca powtarzające się współwystąpienia wyrazów, 3) *Ścieżki* – podanie częstości n-gramów wyrazowych⁷, tworzących kolokacje wraz z wyrazami występującymi między nimi. Warto zaznaczyć, że ośrodkami kolokacji mogą być również jednostki wielowyrazowe, w przypadku których możliwość jednoczesnego zachowania różnych pozycji kolokatów daje dostęp do wyekscerpowania dłuższych wariantów i połączeń wyrazowych dla wybranej jednostki. Tego typu rozwiązania techniczne pozwalają odnaleźć kolejno występujące połączenia dla fatyzmu *na marginesie*, którego wyniki wyszukiwania przedstawiają się następująco (przy zaznaczeniu przysłówków i czasowników jako pożądanых części mowy): *na marginesie wspomnę; na marginesie pragnę zauważyć; na marginesie zapytam; na marginesie przyznam szczerze* itp.

3.7. Narodowy Korpus Języka Rosyjskiego

3.7.1 Ogólna Charakterystyka NKJR

Narodowy Korpus Języka Rosyjskiego (NKJR) dostępny na stronie internetowej www.ruscorpora.ru, powstał pod kierunkiem W. W. Winogradowa w Instytucie Języka Rosyjskiego Rosyjskiej Akademii Nauk. Zawiera on teksty od połowy XVIII do początków XXI w., tj. zarówno starsze, jak i współczesne literackie, ustne, gwarowe teksty rosyjskie (Duszkin 2014: 144). W korpusie znajdują się teksty oryginalne (nieprzetłumaczone), głównie dzieła literackie (proza, dramat i poezja), ale także eseje publicystyczne, prace popularno-naukowe, zapisy oficjalnych przemówień, korespondencja, pamiętniki, dokumenty itp. Obecnie NKJR składa się z następujących podkorpusów:

- **korpus główny** (*основной корпус*), w którym zawarte są głównie teksty prozatorskie XVIII–XXI w. Jest korpusem zrównoważonym z uwagi na zachowane proporcje ilościowe pomiędzy gatunkami tekstów, takimi jak publicystyka, teksty naukowe, literatura religijna itp.
- **korpus składniowy** (*синтаксический корпус*), który dla każdego fragmentu tekstu zawiera opis morfologiczny i strukturę składniową zdania w postaci drzewa zależności, którego węzły zawierają komponenty zdania, a gałęzie oznaczone są nazwami relacji składniowych.

⁷ Częstości ciągów słowoform.

- **korpus prasowy** (*газетный корпус*), który zawiera artykuły z lat 1990–2000, zawierające się w kolejnych podkorpusach: **Korpusie współczesnych mediów** (*корпус современных СМИ*) i **Korpusie mediów regionalnych** (*корпус региональных СМИ*).
- **korpus równoległy** (*параллельный корпус*), który umożliwia wyszukiwanie ekwiwalentu przekładowego dla obranej jednostki wyjściowej z języka rosyjskiego na inny język i odwrotnie. Dostępne są następujące języki: angielski, białoruski, bułgarski, buriacki, hiszpański, włoski, łotewski, niemiecki, polski, ukraiński, francuski, szwedzki, estoński.
- **korpus dydaktyczny** (*обучающий корпус*) stworzony został głównie dla celów dydaktycznych. Składa się z lektur szkolnych i został wyposażony w anotację gramatyczną, odpowiadającą tradycyjnej szkolnej gramatyce.
- **korpus tekstów dialektalnych** (*диалектный корпус*) zawiera zapisy dialektów z różnych regionów Rosji, a także umożliwia szczegółowe wyszukiwanie z uwzględnieniem specyfiki morfologii dialektów.
- **korpus tekstów poetyckich** (*поэтический корпус*) umożliwia nie tylko wyszukiwanie cech leksykalnych i gramatycznych jednostek, ale uwzględnia w zapytaniu konstrukcję wiersza (np. rodzaj rymów).
- **korpus tekstów mówionych** (*устный корпус*) zawiera transkrypcje nagrań tekstów publicznych i niepublicznych, a także dialogów filmowych.
- **korpus akcentologiczny** (*акцентологический корпус*) zawiera informacje o miejscu akcentu w określonej jednostce językowej i uwzględnia akcent w kryteriach wyszukiwania.
- **korpus multimedialny** (*мультимедийный корпус*) zawiera fragmenty filmów z lat 1930–2000, umożliwia wyszukiwanie według form słownych, a także uwzględnia anotację typów gestów i aktów mowy, np. kiwnięcie, zachwyt, ironia, radość itp.
- **korpus MultiPark** (*МультиПАРК - мультимедийный параллельный корпус*) jest pilotażową wersją korpusu multimedialnego, dostępną od 2015 roku i zawierającą dwie produkcje teatralne i adaptację filmową sztuki Gogoła *Rewizor*.
- **korpus historyczny** (*исторический корпус*) zawiera starsze teksty, podzielone na cztery podkorpusy: (1) cerkiewno-słowiański; (2) staroruski; (3) epigrafika; (4) gramoty na brzoźowej korze.
- **korpus panchroniczny** (*панхронический корпус*), który przedstawia proces historyczny kształtowania się jednostki i jej zmiany w strukturze gramatycznej na przestrzeni kilku stuleci.
- **korpus klasyki rosyjskiej** (*Русская классика*) zawiera teksty twórców rosyjskiego języka literackiego od XIX do XX w.

- **korpus „od 2 do 15”** (*Корпус «От 2 до 15»*) zawiera współczesne teksty literackie w języku rosyjskim przeznaczone dla dzieci i młodzieży, opatrzone informacją o preferowanym wieku odbiorców.

- **korpus sieci społecznościowych** (*корпус социальных сетей*) zawiera teksty z blogów, komunikatorów i portali społecznościowych, opatrzone anotacją gatunkową oraz oznaczeniem elementów pozawerbalnych.

Poniżej przedstawiono dane ilościowe dotyczące wielkości poszczególnych podkorpusów NKJR:

Korpus	Liczba tekstów	Число предложений	Число словоупотреблений	% словоупотреблений
Główny	131 488	30 499 014	374 449 975	17,8%
Składniowy	1 304	107 129	1 529 501	0,1 %
Publikacji prasowych	2 717 819	54 260 068	788 969 378	37,6%
Równoległy	6 722	13 100 414	173 288 365	8,3%
Dydaktyczny	1 247	1 184 926	13 761 608	0,7%
Tekstów dialektalnych	2 014	-	599 258	0,0%
Tekstów poetyckich	96 702	1 208 251	13 404 836	0,6%
Tekstów mówionych	4 396	1 965 934	13 984 358	0,7%
Akcentologiczny	1 335 740	13 393 954	134 239 098	6,4%
Multimedialny	1 364	1 009 463	5 696 026	0,3%
MultiPARK	51	79 672	458 531	0,0%
Historyczny	10 424	798 920	14 899 897	0,7%
Panchroniczny	140 326	30 944 770	383 815 697	18,3%
Klasyka Rosyjska	25 882	1 513 674	17 549 885	0,8%
„od 2 do 15”	75	431 193	4 419 420	0,2%
Sieci społecznościowych	1 753 274	14 197 458	157 921 573	7,5%

3.7.2 Wyszukiwanie w NKJR

W zależności od rodzaju zapytania wyszukiwarka NKJR dysponuje trzema podstawowymi narzędziami przeszukiwania baz tekstowych:

- 1) *Поиск точных форм;*
- 2) *Поиск коллокаций;*
- 3) *Лексико-грамматический поиск.*

Pierwsze z nich umożliwia odnalezienie dokładnie sprecyzowanej przez użytkownika formy wyrazowej, przy czym postać fleksyjna wyrazu czy liczba elementów frazy pozostaje dowolna. Należy zauważyć, iż w tym trybie system nie rozróżnia pisowni małą i dużą literą wpisanej jednostki bądź frazy, a ponadto ułatwia sprecyzowanie zapytania poprzez zastąpienie dowolnej liczby liter w słowie symbolem (*). Wyniki wyszukiwania udostępnione zostają w postaci listy konkordancji lub w formie KWIC, a wyposażenie korpusu w dodatkowe instrumenty, takie jak *график по годам*, *частотность*, *статистика* otwiera możliwości dla dokładniejszego zanalizowania jednostki przy korzystaniu z przypisanych jej metainformacji. Poniżej przedstawiono formularz narzędzia *Поиск точных форм* wraz z przykładem konkordancji dla wyrażenia *в раннем детстве*.

Поиск точных форм ? ^

Слово или фраза  ?

в раннем детстве

Искать ▾

Конкорданс KWIC График по годам Статистика Частотность 3-граммы 4-граммы 5-граммы   Скачать ▾ ?

1. Вирусные гепатиты (материалы доложены на 8-м съезде Всероссийского общества эпидемиологов, микробиологов и паразитологов) // «Вопросы вирусологии», 2002.12.02 ↵

Высокая способность вируса вызывать хронизацию процесса с дальнейшим развитием цирроза и первичной гепатокарциномы по крайней мере вдвое увеличивает число смертей от острого гепатита В. Цирроз и гепатокарцинома особенно часто развиваются у тех, кто заразился при рождении и *в раннем детстве*.  ↔

Jedną z najnowszych funkcji korpusu jest możliwość korzystania z kolokatora (*Поиск коллокаций*), który służy do odnalezienia najczęstszych połączeń elementów dla wybranej jednostki wyjściowej:

Ключ

Лемма  ?

добавить условие ▲

Коллокат

Лемма  ? ✕

Грамм. признаки *выбрать* ? ✕

Расстояние

от: до:

добавить условие ▲

Искать

Сбросить

W polu *Ключ* należy określić wyraz, który musi zostać zadany w swojej kanonicznej formie (lemacie) bądź w postaci zestawu cech gramatycznych lub semantycznych jednostki. Nieobligatoryjnym dla wypełnienia polem jest *Коллокат*, który może zostać zastosowany dla obliczenia frekwencji występowania konkretnego połączenia wyrazowego, np. *красная девушка* (w formie lematu: *девушка/красный*). *Грамматические признаки* służą do wskazania pożądanych cech morfologicznych leksemu lub formy wyrazu, np. *существительное/мужского рода/в дательном падеже*. Z kolei funkcja *Расстояние* określa odległość pomiędzy kluczem a kolokatem i nie może wynosić więcej niż pięć słów, np. dla wyszukania wszystkich występujących form określających dla leksemu *завод* należy zaznaczyć parametr *прилагательное* i ustawić odległość od -1 do -1. Wyniki wyszukiwania są uporządkowane zgodnie z frekwencją występowania kolokacji w tekstach wraz z uwzględnieniem pozostałych parametrów i odsyłaczy do przykładów konkordancji, np.:

Коллокации (?)

   Скачать (?)

Ключ	Коллокат	Совместная частота	Частота ключа	Частота коллоката	LogDice ▼	Loglikelihood ▼	MI ³ ▼	t-score ▼	Агр. мера ▼	Конкорданс
завод	металлургический	625	61637	2673	10.06	7989.53	20.13	24.98	22.43	Примеры
завод	кирпичный	574	61637	5453	9.93	6338.44	19.17	23.92	21.41	Примеры
завод	машиностроительный	482	61637	1107	9.82	6886.19	20.24	21.95	21.07	Примеры
завод	сахарный	437	61637	4125	9.68	4830.54	18.63	20.87	19.72	Примеры
завод	автомобильный	440	61637	7101	9.64	4372.21	18.10	20.92	19.46	Примеры
завод	винокуренный	326	61637	462	9.44	5121.99	19.94	18.05	18.97	Примеры
завод	алюминиевый	298	61637	2944	9.31	3264.77	17.82	17.23	17.52	Примеры

Ostatnie narzędzie *Лексико-грамматический поиск* jest wyposażone w najbardziej rozbudowany repertuar dodatkowych parametrów i służy dla określenia sekwencji leksemów lub form wyrazów o określonych cechach gramatyczno-leksykalnych:

Лексико-грамматический поиск ? ^

По умолчанию поиск ведется по предпочтительным разборам, слова могут совпадать.

Поиск учитывает:

- ☒ все разборы
- ☐ предпочтительные разборы

При поиске:

- ☒ совпадения слов исключаются
- ☐ слова могут совпадать

Слово 1 +

Лемма ? ×

Словоформа ? ×

Грамм. признаки **выбрать** ? ×

Семантика основная **выбрать** ? ×

☒ 1-е знач. ☒ др. знач.

Синтаксические признаки **выбрать** ? ×

Доп. признаки **выбрать** ? ×

добавить условие ▲

Do niewystępujących w poprzednich narzędziach funkcji należy m.in. wskaźnik *семантика основная*, który umożliwia wybór określonych cech semantycznych i słowotwórczych wyrazu, np. *транспортные средства*, *положительная оценка*, *диминутивы*. Wskazanie charakterystyki składniowej umożliwia pole *Синтаксические признаки* z przykładowymi opcjami parametrów, takimi jak *подлежащее*, *обращение*, *предлог*, a korzystanie z cech określonych w polu *Дополнительные признаки* umożliwia wyszukiwanie słowa w określonej pozycji: przed i po znaku interpunkcyjnym, na początku i na końcu zdania.

3.7.3 Rodzaje anotacji w NKJR

Proces oznaczenia form wyrazowych w NKJR przebiegał w oparciu o tradycyjną gramatykę słownikową języka rosyjskiego, co spowodowało ujednolicenie stosowanej nomenklatury i pozwoliło uniknąć sporów terminologicznych. W korpusie mamy do czynienia z dwoma typami anotowanych tekstów, które stanowią odrębne dwa podkorpusy – z formami homonicznymi i bez nich. Anotacja pierwszego wykorzystuje program *mystem* firmy Yandex, który przeprowadza analizę morfologiczną tekstu, jest zdolny redukować słowa do ich formy kanonicznej i konstruować hipotetyczne analizy. Podkorpus pozbawiony homonimii jest anotowany w sposób automatyczny przy pomocy parserów, tj. programów z wbudowanymi słownikami, np. *Грамматический словарь русского языка А.А. Зализняка*.

Struktura *anotacji morfologicznej* dla konkretnej jednostki w NKJR zawiera cztery podstawowe typy informacji:

1. Wskazanie formy lematy, czyli hasła słownikowego i jego przynależności do określonej części mowy,
2. Zbiór cech gramatycznych danego wyrazu (np. rodzaj rzeczownika),
3. Cechy gramatyczne lub fleksyjne leksemu (np. aspekt czasownika),
4. Dodatkowe informacje o niestandardowej formie gramatycznej, zniekształceniach itp.

Znacznik tagów został oparty na systemie łacińskich znaków skróconych (np. *num* – liczebnik, *adv* – przysłówek), których uprzednia znajomość umożliwia skonstruowanie bardziej specjalistycznego zapytania, składającego się wyłącznie ze wskazanych oznaczeń. Schemat *anotacji syntaktycznej* polega na przedstawieniu struktury zdania w postaci drzewa zależności, w którym poszczególne strzałki odchodzą od słowa nadrzędnego (ros. *хозяин*) w kierunku wyrazów zależnych (ros. *слыша*) wraz z oznaczeniem związku syntaktycznego. Należy zaznaczyć, że prezentowane w korpusie znaczniki syntaktyczne występują w formatach SinTagRus i Universal Dependencies, które znacząco różnią się nazewnictwem znaczników i zasadami określania relacji pomiędzy elementami konstrukcji zdaniowych. Wysoko rozwinięty i udoskonalony system znakowania w korpusie rosyjskim przyczynił się do możliwości wprowadzenia *anotacji semantycznej*, która odbywa się w sposób automatyczny przy użyciu programu Semmarkup. Znaczniki semantyczne zostały oparte na systemie klasyfikacji słownictwa rosyjskiego przyjętym w bazie danych Lexicograph, a ich repertuar pozostaje zależny od oznakowanej części mowy wyrazu. Dodatkowe informacje leksykalno-semantyczne pogrupowane zostały w następujące grupy: 1. Taksonomia (klasa tematyczna leksemów), 2. Mereologia (dotycząca relacji „całość”- „część”), 3. Topologia (stan topologiczny obiektu), 4. Ocena (pozytywna i negatywna), 5. Aspekt słowotwórczy (np. zdrobnienia i zgrubienia).

3.8. Narodowy Korpus Języka Czeskiego

3.8.1 Ogólna Charakterystyka

Narodowy Korpus Języka Czeskiego (NKJCz) należy do największych korpusów języków słowiańskich i pomyślany jest jako projekt w kontynuacji, którego poszczególne podkorpusy odwzorowują i monitorują różne odmiany języka czeskiego, udostępniając odpowiednie narzędzia do obserwacji danych tekstowych. Powstanie korpusu datuje się na rok 1994 z inicjatywy wieloletniego dyrektora projektu Františka Čermáka, a za jego realizację

odpowiada Instytut Czeskiego Korpusu Narodowego na Uniwersytecie Karola w Pradze (Čermák 2017: 69). Konieczność stworzenia obszernego i reprezentatywnego korpusu czeszczyzny wynikała z potrzeby posiadania solidnej bazy materiałowej dla tworzenia nowych i jakościowo lepszych słowników jednojęzycznych, podręczników do nauki gramatyki czy innych języków obcych. Integralną częścią pracy NKJCz jest tworzenie własnej metodologii i rozpowszechnianie dziedziny lingwistyki korpusowej w obrębie językoznawstwa matematycznego. Z uwagi na niekomercyjny charakter projektu dostęp do korpusów jest bezpłatny, a korzystanie z pełnej wersji jest możliwe po zarejestrowaniu się na stronie www.korpus.cz. Ogólny zakres przetworzonych tekstów przekracza 4,5 miliarda słów w korpusach jednojęzycznych czeskich oraz 1,5 miliarda słów w korpusach równoległych. Bazę tekstową tworzy w tym momencie 32 korpusy języka pisanego (synchroniczne), 14 korpusów języka mówionego, jeden korpus diachroniczny i kilkadziesiąt korpusów równoległych i obcojęzycznych. Poniżej przedstawiono charakterystykę przykładowych korpusów NKJCz dla wybranych kategorii:

Korpusy języka pisanego

	Liczba słów	Lematyzacja	Anotacja morfologiczna	Rok wydania	Charakterystyka
SYN (11 wersja)	5 mld.	Tak	Tak	2010	Łączy synchroniczne korpusy pisane serii SYN
ONLINE (2. generacja)	>6 mld.	Tak	Tak	2020	korpus monitorujący czeski Internet
JEROME	95 mln	Tak	Tak	2013	jednojęzyczny, porównywalny korpus do badań nad przetłumaczonymi tekstami czeskimi
Koditex	10,8 mln	Tak	Tak	2018	korpus stworzony na potrzeby wielowymiarowej analizy języka czeskiego

Korpusy języka mówionego

	Liczba słów	Lematyzacja	Anotacja morfologiczna	Rok wydania	Charakterystyka
ORAL (1 wersja)	5,4 mln	Tak	Tak	2017	referencyjny zbiór mówionego języka czeskiego (obejmuje Czechy, Morawy i Śląsk)
ORATOR (2 wersja)	1,2 mln	Tak	Tak	2019	Korpus referencyjny monologu z transkrypcją jednopoziomową
PMK	675 tys.	Nie	Nie	2001	Praski korpus mówiony. Transkrypcja nagrań przemówień praskich z lat 90. XX wieku

Korpusy języków obcych

	Liczba słów	Lematyzacja	Anotacja morfologiczna	Rok wydania	Charakterystyka
InterCorp	5,3 mld.	Tak	Tak	2008	wersjonowany korpus równoległy zawierający oprócz czeskiego 61 innych języków
CODIT	27 mln.	Nie	Nie	2021	diachroniczny korpus języka włoskiego obejmujący okres od XIII wieku do 1947 roku
Aranea	1000mln.	Tak	Tak	2014	porównywalne korpusy internetowe dla różnych języków

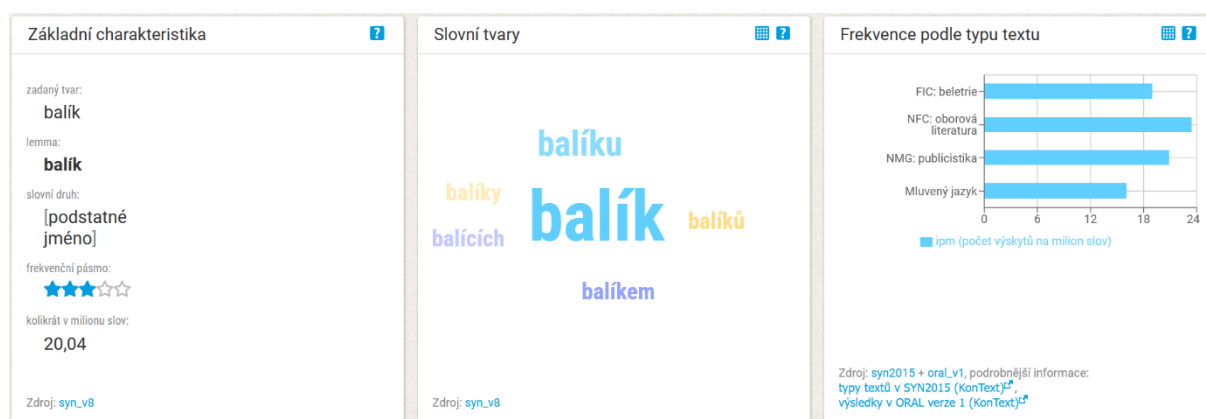
3.8.2 Wyszukiwanie w NKJCz

NKJCz udostępnia szereg aplikacji wyszukujących, z których każda jest wyposażona w dodatkowe narzędzia, umożliwiające przeprowadzenie konkretnego rodzaju analizy. Podstawowym programem do przeszukiwania baz korpusowych jest aplikacja *KonText*, która pozwala na tworzenie prostych i złożonych zapytań, obliczanie rozkładu częstości czy obserwację kolokacji. Poszczególne funkcje dotyczą m.in. wyboru określonego podkorpusu, gatunku tekstu, liczby wyrazów kontekstowych czy pożądaných części mowy, które mają wystąpić w poświadczeniach. Typowym wynikiem wyszukiwania jest lista konkordancji w formacie KWIC, zawierająca wszystkie słowa lub wyrażenia pasujące do zapytania wraz z dodatkowymi metainformacjami, takimi jak bibliografia, płeć autora czy rok publikacji. Pozyskane dane można sortować według zakresu kontekstu, frekwencji formy jednostki czy

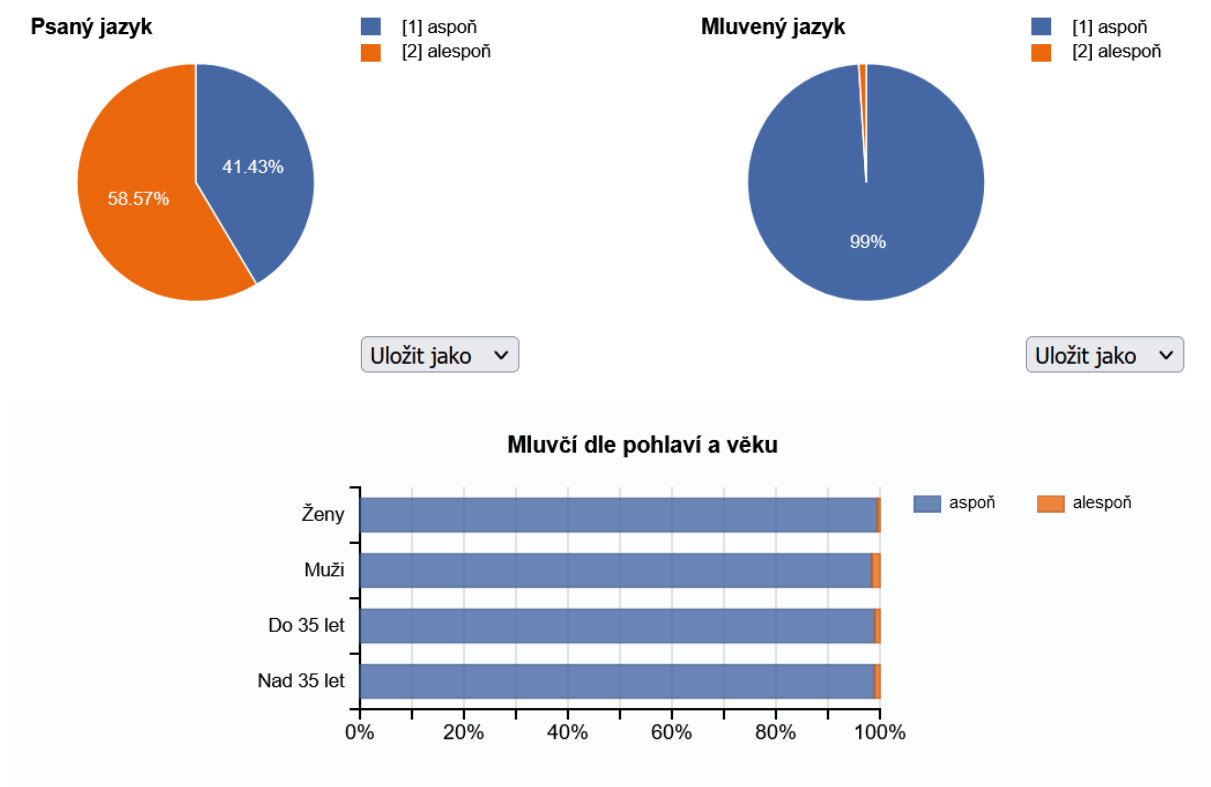
filtru pozytywnego i negatywnego, który umożliwi pozostawienie bądź usunięcie konkordancji zgodnie z określonym warunkiem. Poniżej przedstawiono zrzut ekranu i wyniki wyszukiwania dla zapytania *To je jedno*:

Výskytů: 10 591 i.p.m.: 20,44 (vztaženo k celému korpusu) ARF: 5 418,89 Výsledek je seřazen			1 / 265
Výběr řádků: základní			
☐	abele-ostice	z jedné určité zahrady . (Posmutní .) Ale	to je jedno – stejně se to pořádně nedá jíst ... V tomto
☐	abe-pisecna_zena	, jídlo , taky zapomněl vzít to cyankáli ... Ale	to je jedno , žena stejně těžko přijde na tu skryš ... Tiše
☐	ackroyd-londyn	vedle sebe , jedno bez druhého si nedovedeme představit .	to je jedno z tajemství moci města . Tyto zbytky minulosti jsou teď
☐	adams-daleka_cesta	z kolonie ? " . Ano . Co nejdřív .	to je jedno kam . " , Jen ty a já ? "
☐	adams-daleka_cesta	své nože ale není , " upozornil Žalud .	to je jedno , " řekl Hlavoun , ale hned ho napadlo :
☐	adams-restaurant_na_ko	. " Obchodní dům Harrods zničili Vagoni . " "	to je jedno , " " Říkáš ty . " " Poslouchej .
☐	adams-restaurant_na_ko	. Jemnej bílej písek , respektive cukr . Cokoliv .	to je jedno . Cukr je taky dobrej . A když je plná
☐	adams-slohem_a_dik	A není tak legrační jako ta tvoje , ale ...	to je jedno . " Pohlédla dolů . Arthur přímo hmatatelně cítil ,
☐	adams-zivot_vesmír	dodal . " Tady stojím , mozek velikosti ... ale	to je jedno . " Několik vteřin se zuřivě ploužil dál . "
☐	adams-zivot_vesmír	přes hubu ! " " Hm hmhm ? " "	to je jedno přes kterou . Teď mi něco řekni . Jedno zahuhňání
☐	adichie-amerikana	na tom Emenikem nezdá , " prohlásila . , Ale	to je jedno , hlavně aby zmizel , když se budeme zabývat stropem
☐	adichie-amerikana	že to , je hrozně komplikovaná záležitost " . Ale	to je jedno , stejně radši držte klapáčku , zvlášť když od dotyčného
☐	adler-slozka64	dostalo do zpomaleného záběru . Rita potřásla hlavou . "	to je jedno . Máš aspoň mlíko , abych si ho tam trochu
☐	adler-slozka64	odchodu . Netě se rozhlédla kolem sebe . Někte	to je jedno kde , budou problémy muset začít . Otázka zněla kde
☐	adler-zabijaci	" Co bychom měli ? " " Ale nic ,	to je jedno . Mám pro tebe další úkol , a sice že

Aplikace *Slovo v kostce* slouží do tvorby základního přehledu způsobu užívání daného slova a obsahuje informace, pocházející ze zdrojů NKJCz. Okrom možnosti vyhledání slova v češtině (Vyhledat slovo) dostupná je rovněž možnost odnalezení jeho překladů na jiné jazyky (Hledat ve dvou jazycích/přeložit) v souboru mnohojazyčného korpusu InterCorp. Výsledky automatické analýzy provedené jednotkou jsou prezentovány v podobě karet, které obsahují následující informace: *Základní charakteristika*; *Slovní tvary*; *Frekvence podle typu textu*; *Kolokace*; *Ukázky kolokací v textech*; *Frekvence výskytu slova v čase*; *Podobně používaná slova*; *Frekvence slova podle mluvčích*; *Oblasti podle tradičního nářečního členění*; *Ukázka autentické promluvy*. Jednou z základních funkcí je rovněž možnost porovnání profilů dvou nebo více slov nebo frází v modulu *Porovnat slova*, a přesnost a spolehlivost získaných výsledků závisí na typu použité anotace. Výsledky vyhledání pro jednotku *balík* jsou prezentovány níže s ohledem na tři zmíněné charakteristiky:



Narzędzie *SyD (Korpusový průzkum variant)* jest przeznaczone do badań konkurencyjnych wariantów jednostek, występujących zarówno we współczesnym języku czeskim, jak i starszej czeszczyźnie. Aplikacja składa się z dwóch odrębnych części: 1) *synchronicznej*, która porównuje warianty ze względu na ich częstość występowania w tekstach, 2) *diachronicznej*, umożliwiającej prześledzenie tendencji rozwojowych określonych wariantów. Część synchroniczna dostarcza również informacji o rozmieszczeniu wyrazu w tekstach pisanych oraz w języku mówionym (uwzględnia informacje o płci, wieku, wykształceniu i przynależności regionalnej autora wypowiedzi). Z kolei analiza diachroniczna bazuje na korpusie Diakorp i nie uwzględnia opcji lematyzacji jednostek wyjściowych z powodu braku ich ręcznej weryfikacji. Zapytania można formułować poprzez dokładne wpisanie dwóch jednostek bądź przy użyciu języka zapytań CQL, a wyniki wyszukiwania prezentowane są w postaci tabel i wykresów. Konkurencja może dotyczyć różnicy w pisowni, morfologii, aspektu syntaktycznego czy szyku wyrazów i synonimów, a do analizy wykorzystywane zostają dane z korpusów mówionych i pisanych NKJCz. Wpisanie zapytania *aspoň/alespoň* w korpusie synchronicznym udostępnia następujące przykładowe informacje:



Do empirycznej interpretacji tekstów i ich porównania z zasobami korpusowymi służy aplikacja *KWords*, której celem jest identyfikacja słów kluczowych, wykazujących wyższą frekwencję użycia w badanym tekście niż w korpusie referencyjnym. Wyekscerpowane przez

narzędzie jednostki stanowią podstawę do dalszej analizy tekstu, któremu przypisywana jest również jego koncentracja tematyczna. Wyniki analizy słów kluczowych są nierozdzielnie związane z wyborem korpusu referencyjnego, który stanowi „neutralne tło językowe” i służy jako fundament dla przeprowadzenia badań komparatystycznych. Wprowadzony przez użytkownika tekst przechodzi proces lematyzacji i tokenizacji wraz z obliczeniem częstości występowania wszystkich słów w badanym tekście. Wyniki analizy zawierają listę zweryfikowanych słów kluczowych wraz z informacją o ich frekwencji, a także podstawowe dane o załadowanym fragmencie tekstu, wykres rozproszenia jednostek i analizę słów kluczowych. Dla zilustrowania pracy narzędzia została wykorzystana część pracy doktorskiej pt. *Národní parky z pohledu české právní úpravy* (Konečná 2017/2018), której analiza dostarczyła następujących informacji:

Klíčová slova				
Tematická koncentrace				
Atributy klíčových slov		Relativní frekvence		Stat. hodnota
word		text	ref.	
naplňování		12875.5	2.9	36.1
parků		17167.4	5.9	47.9
zóna		17167.4	6.4	47.3
národních		17167.4	16.4	40.8
péče		21459.2	72.7	41.4
cíle		12875.5	45.2	22.4
území		12875.5	109.4	18.0

Klíčová slova			
Tematická koncentrace			
Atributy klíčových slov		Pořadí ▲	Frekvence
word	pos		
péče	N	5	5

Aplikacja *Treq* jest narzędziem służącym do wyszukiwania odpowiedników tłumaczeniowych, z którego można korzystać od września 2014 roku, a zawarta w nim baza tekstowa czerpie z zasobów InterCorp. Możliwym jest zadanie jednowyrazowego lub wielowyrazowego zapytania i jego ograniczenia wyszukiwania jednostki w określonych kolekcjach (np. Biblii, napisów filmowych, literatury pięknej). Znaczące zmiany przyniosła wersja 2.1 aplikacji *Treq*, w której możliwe stało się zastosowanie wyrażeń regularnych (*RegEx*) i użycie parametru (*A=a*), dzięki któremu można wprowadzić zapytanie bez rozróżniania wielkich i małych liter. Aktualnie aplikacja zawiera bazę ekwiwalentów tekstowych dla czterdziestu dwóch języków, z których czeski, angielski i hiszpański są językami obowiązkowymi podczas tworzenia zapytania. Należy zaznaczyć, iż ekwiwalenty oferowane przez aplikację można uznać za potencjalne odpowiedniki przekładowe poprzez wygenerowanie przez narzędzie listy jednostek „pretendujących” do bycia jednostkami przekładowymi. Wyniki zapytania zostają automatycznie posortowane według liczby wystąpień z uwzględnieniem udziału procentowego w całkowitej objętości korpusu InterCorp.

Treść dla zapytania *pączek* w przekładzie polsko-czeskim z zaznaczeniem funkcji *lemmaty* (umożliwiającej wyszukanie wszystkich form fleksyjnych danego leksemu) przynosi następujące informacje:

The screenshot shows the NKJCz search interface. At the top, there are dropdown menus for 'Výchozí jazyk' (Polština) and 'Cílový jazyk' (Čeština), a 'Omezit na' (Kolekce: 6) dropdown, and a search input field containing 'pączek'. Below the search bar are checkboxes for 'Lemmata' (checked), 'Viceslovné', 'Regulární', and 'A = a' (checked). A yellow 'Hledej' button is to the right of the search bar. Below the interface is a table with four columns: 'Frekvence', 'Procenta', 'Polština', and 'Čeština'.

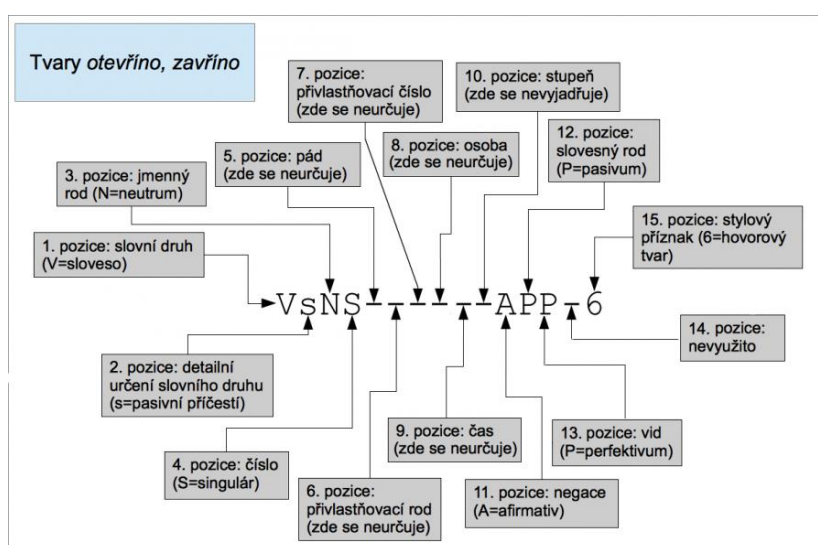
▲ Frekvence ▼	▲ Procenta ▼	▲ Polština ▼	▲ Čeština ▼
212	58.2	pączek	kobilha
19	5.2	pączek	poupě
12	3.3	pączek	pupen
8	2.2	pączek	kobilžek
7	1.9	pączek	kompot
6	1.6	pączek	kobilžka
5	1.4	pączek	kolač

3.8.3 Anotacja w NKJCz

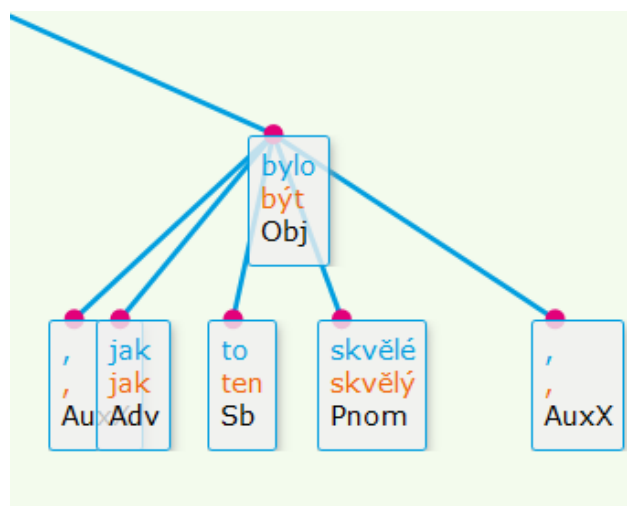
System anotacji zewnętrznej i wewnętrznej w NKJCz pozostaje kwestią zróżnicowaną z uwagi na niejednorodny zakres stosowania anotacji w poszczególnych korpusach. Korpus SYN verze 11 udostępnia szereg szczegółowych informacji, dotyczących zarówno danych o autorze, wydawnictwie i roku publikacji, jak i charakterystyk morfologicznych konkretnych jednostek. Nowatorskim rozwiązaniem jest wykorzystanie pilotażowej *anotacji frazowej*, która w sposób automatyczny wyszukuje i identyfikuje zarejestrowane kolokacje i idiomy, dla których źródłem stał się *Český slovník frazeologie a idiomatiky* Františka Čermáka. Metody dochodzenia do jednostek wielowyrazowych w korpusach opierają się na możliwości wskazania konkretnych właściwości gramatycznych poszczególnych form wyrazowych czy ustalenia kolejności wyrazów w konkretnej frazie. Zebrane w korpusie *KSK-DOPISY* teksty korespondencyjne opatrzone zostały m.in. sygnaturami listów, informacjami o relacjach łączących nadawcę z adresatem tekstu czy danymi lokalizacyjnymi, dotyczącymi pochodzenia obu uczestników wymienianych wiadomości. Z kolei specjalistyczny korpus języka mówionego BMK uwzględnia przede wszystkim cztery kryteria socjolingwistyczne, do których należą: płeć [M (*mężczyzna*) i Z (*kobieta*)], wiek mówców [I (*młodzi: ok. 20–35 lat*) oraz V (*starsi: 35 lat i więcej*)], wykształcenie [B (*podstawowe*) a A (*wyższe łącznie ze studentami*)], a także rodzaj wypowiedzi [F (*formalny*) i N (*nieformalny*)] (Hebal-Jezierska 2014: 35).

Należy odnotować, iż na poziomie *anotacji wewnętrznej* proces lematyzacji przybiera charakter dwupoziomowy, tzn. każdej jednostce przypisane zostaje jej hasło podstawowe i

forma sublematy. Podczas gdy lemat może zawierać kilka wariantów tego samego wyrazu (np. lemat *filozofie* przypisywany zostaje do wszystkich jednostek z rdzeniem *filozof* i *filosof*), o tyle sublematy definiują podgrupy form według tej wariantywności (sublemat *filozofie* odnosi się wyłącznie do jednostek z rdzeniem *filozof*, pomijając formę *filosof*). Znaczniki zostały tak skonstruowane, aby każda pozycja odpowiadała określonej kategorii morfologicznej, zgodnie z przyjętą tradycyjną koncepcją językową i składa się z ciągów 15 znaków, do których należą m.in. *slovní druh*; *jmenný rod*; *číslo*; *pád*; *negace*; *detailní určení slovního druhu*; *osoba*; *vid*; *varianta* (*stylový příznak*). Poniżej przedstawiono przykład anotacji morfologicznej dla wyrazów *otevřino*, *zavřino*:



Niektóre z korpusów są wyposażone w system *anotacji syntaktycznej*, uwzględniającej zależności pomiędzy poszczególnymi wyrazami w zdaniu, a ich funkcjami kompozycyjnymi. W tym celu użyty został automatyczny analizator składniowy, tzw. *parser*, który każdemu tokenowi przypisuje inny token (czes. *rodič*), od którego ten pozostaje zależny. Pojedyncze części zdania opatrzone zostały znacznikami składni, które odpowiadają tradycyjnym funkcjom kompozycyjnym, takim jak orzeczenie (Pred), podmiot (Sb), przymiotnik (Atr) itp. Struktura składniowa zdania jest wyświetlana jako drzewo zależności, którego gałęzie oznaczają relacje między wyrazami, uzupełnionymi o informacje dotyczące ich formie podstawowej. Oznakowanie syntaktyczne dla fragmentu zdania *jak to bylo skvělé*, zaprezentowano poniżej:



3.8.4 Kolokacje w NKJCz

NKJCz, w przeciwieństwie do korpusów języka polskiego czy rosyjskiego, nie posiada odrębnego narzędzia kolokacyjnego, lecz integruje możliwość wyszukiwania złożonych kolokacji jako jedną z kluczowych funkcji w ramach własnego zestawu narzędzi analitycznych. W zależności od specyfiki danego korpusu, repertuar dostępnych funkcji wyszukiwujących różni się dla każdego zbioru tekstowego, umożliwiając użytkownikowi formułowanie zapytań za pomocą odpowiednich pól w interfejsie wyszukiwania, który przedstawiono poniżej:

Kolokace ✕

Atribut: word

V kontextu: do

Minimální frekvence kolokátu v korpusu:

Minimální frekvence kolokátu v kontextu:

Kolokační míry:

Zobrazit Použit pro třídění

T-score	☒	☐
MI	☒	☐
MI3	☐	☐
log likelihood	☐	☐
min. sensitivity	☐	☐
logDice	☒	☒
MI.log_f	☐	☐
relative freq.	☐	☐

Vytvořit seznam kolokátů

Wartości w pozycji *Atribut* (która umożliwia precyzyjne określenie właściwości kolokatu) różnią się w zależności od wybranego zbioru tekstowego, jednakże do najczęstszych z nich można odnieść poniższe parametry: 1) *word* (konkretna forma wyrazu), 2) *lemma* (forma słownikowa wyrazu), 3) *tag* (znakowanie morfologiczne) 4) *case* (przypadek w jakim winien występować kolokat) czy 5) *verbtage* (oznaczenie kategorii gramatycznych czasownika) i wiele innych. Z kolei funkcja *V kontextu* czyni możliwym oznaczenie miejsca występowania kolokatu, tj. liczby ujemne oznaczają pozycje poprzedzające KWIC, a dodatnie pozycje następujące po nim. *Minimální frekvence kolokátu v korpusu* pozwala określić minimalną całkowitą częstość występowania jednostki, która może zostać włączona do listy wyekscerpowanych kolokatów, jeśli spełnia założony warunek liczbowy. Zadanie wartości w pozycji *Minimální frekvence kolokátu v kontextu* określa minimalną liczbę występowania kolokatu w zadanej odległości od jednostki podstawowej, którą wyraz kolokacyjny musi spełnić, aby został uwzględnionym przez korpus w listach konkordancyjnych. Ostatnią z opcji są tzw. *kolokační míry* (formuły matematyczne używane do znajdowania kolokacji w korpusie), dzięki którym użytkownik może wybrać, która z dostępnych miary asocjacyjnych będzie zastosowana w procesie poszukiwania kolokatu dla wskazanej jednostki. Wartym podkreślenia wydaje się fakt, iż twórcy korpusu już na wstępie zwracają uwagę na brak jednoznaczności oraz powszechnej zgodności w językoznawstwie w kwestii terminu *kolokacja*. Z tego powodu, w obrębie wspomnianego pojęcia uwzględnili oni zróżnicowane typy połączeń wyrazowych, które zawierają:

- Powszechnie używane kolokacje (*letní šaty, vejce naměkko*);
- Frazemy i idiomy (*ležet ladem, růžové brýle*);
- Terminy wielowyrazowe (*infarkt myokardu, červí díra*);
- Wielowyrazowe nazwy własne (*Andělská Hora, Kostelec nad Černými Lesy*).

Wyszukanie najczęściej używanych połączeń leksykalnych jest możliwe również dzięki zastosowaniu, m.in.: specjalistycznych zapytań CQL (np. funkcja *meet and union*) czy za pomocą filtrów i dystrybucji częstości, czyli wspomnianych statystycznych miar asocjacji. Automatyczne wyszukiwanie kolokacji polega na korelacji frekwencji jednostek wchodzących w kolokację, częstości całej asocjacji i wielkości korpusu. Szczegółowa instrukcja dla wspomnianych metod wyszukiwania kolokacji jest dostępna pod adresem https://wiki.korpus.cz/doku.php/kurz:hledani_kolokaci. W obszarze badań nad wielowyrazowymi i odtwarzalnymi jednostkami języka mówionego szczególnie przydatnym narzędziem, obsługującym korpus SYN, jest instrument *FRANTA* (FRazémová ANotace

a Textová Analýza), který w sposób automatyczny dokonuje identyfikacji stałych połączeń wyrazowych, a jego baza tekstowa opiera się na omawianej już pracy Františka Čermáka *Slovník české frazeologie a idiomatiky*. Z uwagi na fakt dużej zmienności w realizacji jednostek wielowyrazowych w korpusie, narzędzie pozwala zażądać konkretnej interpretacji morfologicznej jednostek, określić ich spójność czy zaakceptować (bądź odrzucić) występujące zmiany w kolejności wyrazów. Wyodrębnione przez korpus jednostki są oznaczone dwoma atrybutami pozycyjnymi, które należy zaznaczyć podczas próby odnalezienia reproduktów wielowyrazowych:

- *col_lemma* (lemma kolokacji): lemma jednostki wielowyrazowej w postaci hasła słownikowego w jego podstawowej formie (mianownik liczby pojedynczej, bezokolicznik itp.); Wyniki poszczególnej formy wyrazów są oddzielone podkreślnikiem i w wielu przypadkach dostarczają gotowych połączeń wyrazowych o wysokiej frekwencji występowania w korpusie, np. dla słowa *pes* zadanie niniejszego atrybutu z odległością -4/5 w funkcji *v kontextu* przyniosło następujące przykładowe jednostki wielowyrazowe: *zakopaný_pes* (*v tom ten zakopaný pes*), *každý_pes_jiná_ves*; *starého_psa_novým_kouskům_nenaučíš* (gotowe reproduktury) itp.
- *col_type* (typ kolokacji): dwuliterowy znacznik, którego pierwsza litera określa typ jednostki wielowyrazowej, a druga służy do rozróżnienia między wyrazami nadrzędnymi (H) i zależnymi (Z). Część frazy oznaczona jako główna (H) nie musi być składniowo lub semantycznie dominująca, główną motywacją dla tego rozróżnienia jest możliwość prawidłowej oceny częstości frazy w korpusie (tylko jedna pozycja frazy jest zawsze oznaczona jako H - zwykle ostatnie słowo), podczas gdy wszystkie inne są oznaczone jako Z. Symbole wraz z przypisanymi im znaczeniami ujęto w następującej tabeli:

1. pozice	význam	příklad viceslovného lemmatu	příklad jeho označení v korpusu
K	neslovesné kolokace	cesta_do_pekla	Dluhy jsou cestou[KZ] do[KZ] pekla[KH].
J	viceslovné spojky	a_právě_proto	A[ZZ] právě[ZZ] proto[ZH] se tak nejspíš nestalo.
M	přísluví, okřídlená rčení, citace apod.	čih_čertu_dobře_peklem_se_tě_odmění	Čih[MZ] čertu[MZ] dobře[MZ], peklem[MZ] se[MZ] tě[MZ] odmění[MH].
P	přirovnání	tma_jako_v_peklu	Uvnitř byla tma[PZ] jako[PZ] v[PZ] peklu[PH].
S	větné výrazy	rozpoutalo_se_hotové_peklo	V tu chvíli se[SZ] rozpoutalo[SZ] hotové[SZ] peklo[SH].
V	slovesné kolokace	mít_z_pekla_štěstí	Měl[VZ] jste skutečně z[VZ] pekla[VZ] štěstí[VH].

Wyniki uwzględnienia niniejszej funkcji w ponownym zapytaniu *pes* przyniosło następujące rezultaty:

1. MH (np. *Znáš to přísloví, že pes, který štěká, nekouše*),
2. MZ (np. *Život je pes, a přitom se neví, jaké bude zítra zvíře*),
3. PZ (np. *Vyskočil a otrěpal se jako pes*) itp.

Jedną z innych możliwości, która wykazuje ogromną przydatność przy analizie wariantywności jednostek wielowyrazowych i w szczególności porównań, jest funkcja zastąpienia słowa w lemacie kolokacyjnym *col_lemma*. W celu zastosowania danej opcji, należy sformułować szczegółowe zapytanie, które zawiera jedną z poniższych formuł zastępczych:

- **A** - zastępuje dowolny przymiotnik w lemacie kolokacji (np. *A_jako_kamén* → *Tuhé jako kamén*).
- **N** - zastępuje dowolny rzeczownik w lemacie kolokacyjnym (np. *N_jako_hrom* → *Návrat jako hrom*)
- **V** - zastępuje dowolny czasownik w lemacie kolokacyjnym (np. *V_jako_zářezaný* → *poslouchal jako žářezaný*)

Automatyczne oznaczanie ustalonych kolokacji w NKJCz nie jest pozbawione niedoskonałości, co odzwierciedla się chociażby w braku uwzględnienia syntaktycznej struktury i walencyjności czasowników, prowadzącymi do ryzyka błędów identyfikacji niektórych złożonych frazeologizmów. Na przykład przy wyszukiwaniu wyrażen typu *mít figuru* (dosł. mieć figurę); *mít náladu* (mieć nastrój) czy *mít právo* (mieć prawo) mogą wystąpić fałszywe pozytywne wyniki, jak w zdaniu: *Kdy jindy by se měly zúročit všechny snahy o dokonalou figuru* (Kiedy indziej miałyby się opłacić wszystkie starania o doskonałą figurę). Dlatego należy mieć na uwadze fakt, że pilotażowa wersja systemu anotacji frazeologicznej nie jest wolna od błędów, które mogą wynikać m.in. z powodu błędów morfologicznych lub braku odpowiednich wyrażen w zasobach korpusowych. Ponadto, automatyczne oznaczanie nie we wszystkich przypadkach jest w stanie rozróżnić znaczenie frazeologiczne od dosłownego, co może być widoczne w wyrażeniach takich jak *lízat si rány* (dosł. „lizać rany”), które może oznaczać dosłowną czynność (np. *Kocour si líže rány* – „Kot liże rany”), ale także metaforycznie, tj. leczyć rany emocjonalne.

ROZDZIAŁ IV

Korpusy wielojęzyczne

Powstanie i rozwój korpusów jednojęzycznych przyczyniły się do podjęcia szeregu badań i specjalistycznych analiz w obszarze wybranego języka, dając tym samym początek nowatorskim opracowaniom naukowym z zakresu m.in.: glottodydaktyki, leksykografii jednojęzycznej czy szeroko pojętej lingwistyce komputerowej. Tworzenie i eksploatacja korpusów jednojęzycznych nadal pozostaje główną domeną współczesnego językoznawstwa korpusowego, jednakże w wyniku systematycznego udoskonalania technik cyfrowych i wzmożonej intensyfikacji prac kontrastywnych zaczęły powstawać zbiory tekstowe, uwzględniające nie jeden, a więcej języków.

Korpusy wielojęzyczne, bo o nich mowa, w swoich zasobach zawierają źródła z wielu języków i posiadają zarówno korpusy porównywalne (teksty w dwóch lub więcej językach, reprezentujące jednakową tematykę), jak i korpusy równoległe, w których tekst oryginalny jest ilustrowany wraz z jego tłumaczeniem. Termin *wielojęzyczny* w wąskim znaczeniu musi oznaczać zbiór tekstowy, uwzględniający co najmniej trzy języki. Biorąc pod uwagę fakt, że przedstawione korpusy są zjawiskiem stosunkowo nowym, istnieją znaczące rozbieżności w kwestii terminologii stosowanej w odniesieniu do tych korpusów (np. uznanie korpusu jednojęzycznego za korpus porównywalny). Niekiedy korpus równoległy bywa nazywany również korpusem translacyjnym, co może wynikać z opozycyjnego przyjęcia sposobu definiowania terminu *„równoległość”*. Na podstawie zgromadzonych tekstowych baz danych można założyć istnienie trzech głównych typów korpusów, zawierających więcej niż jeden język:

- 1) Typ A: *Teksty źródłowe wraz z tłumaczeniami* (np. korpus OPUS),
- 2) Typ B: *Podkorpusy jednojęzyczne*, zaprojektowane przy użyciu podobnych technik próbkowania (np. podkorpusy Narodowego Korpusu Języka Rosyjskiego),
- 3) Typ C: *Kombinacja typów A i B* (np. korpus równoległy InterCorp) (McEnery, Xiao and Tono 2006: 46–47).

Jedną z najtrudniejszych, a zarazem najistotniejszych zadań stawianych korpusom wielojęzycznym, jest procedura wiązania ekwiwalentów w tekstach, która umożliwia wyodrębnianie sugerowanych par przekładowych i prowadzenie dalszych analiz. Niniejszy etap w przypadku korpusów równoległych musi zostać poprzedzony praktyką wyrównywania

tekstu na poziomie zdań, rzadziej fragmentów tekstów, w obrębie których ma być przeprowadzone automatyczne wiązanie. Skuteczne zastosowanie wyrównywania, bądź zrównoleglania wiąże się z prawidłowym użyciem anotacji dla poszczególnych jednostek leksykalnych poprzez przypisanie oznaczeń morfologicznych i syntaktycznych. Strukturalna identyczność tekstów umożliwia zastosowanie prostych i efektywnych algorytmów wiązania ekwiwalentów, opartych na podobieństwie struktury między oryginałem a tłumaczeniem. Jednak w praktyce oryginalna struktura tekstu źródłowego często nie jest w pełni zachowywana w procesie tłumaczenia, co utrudnia bezbłędne działanie algorytmów opierających się wyłącznie na analogii strukturalnej. Naturalne jest zatem, że w przypadku korpusów równoległych, gdzie występuje relacja oryginał – tłumaczenie, stosuje się inne procedury wiązania niż w korpusach porównywalnych, w których taka relacja nie występuje.

Niemniej jednak, udostępnienie wielojęzycznych baz tekstowych znalazło szerokie zastosowanie m.in. w badaniach interkulturowych, przekładoznawstwie, językoznawstwie kontrastywnym czy wreszcie w dydaktyce. Współczesny rozwój oprogramowania komputerowego do tłumaczenia maszynowego zauważalnie wspiera dostęp do danych równoległych, których połączenie z korpusami porównywalnymi stanowi użyteczne źródło do badań nad translacją. Korpusy wielojęzyczne niewątpliwie przyczyniły się również do znaczących innowacji w obszarze leksykologii porównawczej i w praktyce leksykograficznej. Dotychczasowy stan wymienionych dziedzin odznaczał się koncentracją na zestawieniu izolowanych struktur językowych, co w konsekwencji było powodem nieuwzględnienia wszystkich możliwych kontekstów użycia danej jednostki w artykule hasłowym. Zbiór autentycznych środowisk tekstowych określonego wyrażenia umożliwił zaakcentowanie możliwych aspektów funkcjonalnych jednostki wraz z określeniem stopnia frekwencji porównywalnych struktur. W obszarze leksykologii wykorzystanie korpusów okazało się skuteczne przy badaniu łączliwości wyrazów i ustalaniu dla nich kolokatów, co przełożyło się na kompleksowy opis słownikowy jednostki w opracowaniach leksykograficznych. Weryfikacja znaczeń jednostek leksykalnych uprzednio zarejestrowanych w słownikach dwujęzycznych umożliwiła korektę zastosowanych przekładów, zwłaszcza w odniesieniu do idiomów i wyrażen terminologicznych. Z kolei obecność dużej liczby pełnotekstowych równoległych baz danych przygotowała podstawę dla stworzenia automatycznego systemu tworzenia słowników różnego typu, np. frekwencyjnych czy alfabetycznych. Jak można zauważyć, każdy z wymienionych wyżej korpusów jest przeznaczony do innych celów badawczych, ich wykorzystanie przynosi odmienne rezultaty, a poszczególne ich rodzaje charakteryzują się odmienną architekturą i innym systemem anotacji.

4.1 Korpus porównywalny

Jak już zostało wspomniane, korpus porównywalny zawiera teksty w języku źródłowym oraz innym języku, dla których wspólny punkt styczności opiera się na dobraniu źródeł, dotyczących wybranego tematu czy gatunku przy zachowaniu zasad reprezentatywności i użyciu jednolitego operatora próbkowania. Zawarte w nich teksty nie są tłumaczeniami, a jedynie przykładami tekstów danej dziedziny, np. aktów prawnych, instrukcji technicznych czy dokumentów medycznych, których obecność umożliwia porównywanie występujących form językowych czy określenie frekwencji konkretnych zjawisk w rozpatrywanych językach. Tworzenie tego rodzaju korpusów każdorazowo poprzedza wybór kryteriów doboru tekstów, ustalenie zasad ich oznaczania czy regulacja dopuszczalnej objętości danych (Маник 2019: 228–230). Do podstawowych reguł, które należy stosować w momencie budowy korpusu porównywalnego, należy odnieść poniższe kryteria, w ramach których porównywalność może być rozpatrywana:

- 1) Tematyka – teksty powinny posiadać wspólną tematykę zagadnienia,
- 2) Okres powstania – zebrane źródła powinny pochodzić z jednolitych ram czasowych w celu uniknięcia w tekstach m.in.: archaicznych form językowych,
- 3) Styl – teksty winny mieć zachowaną równowagę stylistyczną,
- 4) Typ publikacji – do korpusu należy włączać materiały pochodzące z analogicznych wydawnictw,
- 5) Medium – zawierające się w zasobach teksty powinny być udostępniane podobnymi kanałami przekazu (Lewandowska-Tomaszczyk 2005: 53).

Należy nadmienić, że założona porównywalność tego rodzaju korpusów polega raczej na stosowaniu technik pobierania próbek tekstowych i względnej równowadze, dotyczącej zachowania proporcji, wspólnego gatunku, dziedziny czy okresu powstania tekstów. W przypadku niezachowania wskazanych wymogów taki korpus może stanowić niewiarygodną podstawę do badań kontrastywnych czy translacyjnych. Nie ulega wątpliwości jednak fakt, iż takiego rodzaju zbiory tekstowe pozwalają ustalić różnice typologiczne i kulturowe w obrębie wybranych języków, określić ich cechy wspólne, a także pozyskać z tekstów i wdrażać nowe terminy w praktyce tłumaczeniowej. Prostota tworzenia korpusu porównywalnego w zestawieniu z korpusem równoległym wynika z braku konieczności umieszczenia wiarygodnego tłumaczenia oryginału, a pozyskanie wielojęzycznych tekstów niezależnych dotyczących wybranej tematyki jest zdecydowanie łatwiejsze. Problematicznym może być brak wystarczającej liczby tekstów pewnej dziedziny bądź zawodność automatycznych

algorytmów w kwestii dostrzegania podobieństw między tekstami i ekscerpcji ekwiwalentów przekładowych.

Stosowane w przypadku korpusów porównywalnych algorytmy przetwarzania i wiązania tekstów nie mogą być oparte na identyfikacji i odpowiedniości poszczególnych elementów struktury jak w przypadku korpusów równoległych. Orientacja w stronę niższego poziomu złożoności, czyli *podejścia leksykalnego* zakłada prawdopodobieństwo wystąpienia w tekstach jednostek języka o zbliżonych cechach morfologiczno-syntaktycznych (Haruno, Ikehara and Yamazaki 1996: 525–530), np. pol. *trawa* – czes. *tráva*. Procedura wiązania ekwiwalentów w korpusie porównywalnym składa się z dwóch głównych etapów: automatycznego wyodrębnienia zakresu obejmującego słownictwo i frazeologię oraz właściwego procesu wiązania ekwiwalentów. Identyfikacja słownictwa opiera się na technikach, takich jak analiza kluczowości wyrazów, która oblicza statystyczną częstość występowania jednostek i porównuje ją z ich frekwencją w referencyjnym korpusie języka ogólnego. W ten sposób wyodrębniane są te leksemy, które w badanych tekstach występują znacznie częściej niż w korpusie referencyjnym. W celu eliminacji szumu generowanego przez wyrazy funkcyjne, stosuje się tzw. listy wykluczeń, które można automatycznie wygenerować na podstawie analizy frekwencji wyrazów w korpusie referencyjnym. Po wyodrębnieniu uzyskuje się listę słownictwa stanowiącego podstawową terminologię danej dziedziny, która następnie służy do tworzenia list jednostek leksykalnych zawierających kluczowe terminy. Następnie, przy użyciu algorytmu automatycznego wiązania ekwiwalentów, zestawiane są ze sobą terminy o podobnym współczynniku kluczowości i zbliżonej pozycji na listach frekwencyjnych w obu językach. Kluczowym elementem procesu są również okna porównań, wyznaczone na podstawie różnic w pozycjach wyrazów na listach, co umożliwia wiarygodną identyfikację ekwiwalentnych jednostek leksykalnych w porównywanych językach (Lewandowska-Tomaszczyk 2005: 59).

Korpusy porównywalne nadal stanowią mniejszość w językoznawstwie korpusowym, a poszczególne zbiory tekstowe są w znacznej mierze tworzone do prywatnych badań naukowych i nie są udostępniane publicznie. Jednym z największych korpusów porównywalnych pozostaje baza ICE (International Corpus of English), który zawiera teksty reprezentujące odmiany języka angielskiego dla celów badań różnic strukturalnych. Zdecydowana większość tekstów pochodzi z transkrypcji języka mówionego i uwzględnia źródła tekstowe zaczerpnięte z ponad dwudziestu krajów, w których angielszczyzna jest językiem urzędowym. Przykładem innego jednojęzycznego korpusu porównywalnego jest zbiór tekstowy JEROME, udostępniony w ramach NKJCz. Niniejsza baza z jednej strony zawiera teksty przetłumaczone na język czeski

z innych języków, a z drugiej – oryginalne i nieprzetłumaczone teksty czeskie. Korpus przeznaczony jest przede wszystkim dla tłumaczy i lingwistów w celu zbadania metod translacyjnych bądź zastosowanych w przekładzie uniwersaliów tłumaczeniowych. Za wielojęzyczny zbiór korpusów porównywalnych można uznać projekt MULTITEXT-EAST, który zawiera całe powieści oraz artykuły publicystyczne w dwudziestu językach słowiańskich, w tym dla polskiego, czeskiego i rosyjskiego.

4.2 Korpus równoległy

Przez *korpus równoległy* należy rozumieć zbiór tekstów oryginalnych w języku L1 wraz z ich tłumaczeniami na jeden lub więcej języków L2. Niniejszy korpus charakteryzuje się dodatkową obecnością oznaczeń metatekstowych (np. morfologicznych) i włączeniem do takiego korpusu pełnych tekstów, a nie wyłącznie reprezentatywnych fragmentów. W ramach danego pojęcia należy wyróżnić korpus równoległy dwujęzyczny, obejmujący wyłącznie zasoby dwóch konkretnych języków bądź wielojęzyczny, reprezentujący w swoich zasobach teksty z więcej niż dwóch języków. W zależności od kierunku tłumaczenia można wyróżnić:

- 1) *korpus jednokierunkowy* (uni-directional corpus), gdy tekst oryginalny jest tłumaczony na inny język,
- 2) *korpus dwukierunkowy* (bi-directional corpus), zawierający przełożone teksty języka wyjściowego i zwrotne tłumaczenia przekładów,
- 3) *korpus wielokierunkowy* (multi-directional corpus), zawierający tłumaczenie tekstu źródłowego na wiele języków (McEnery, Xiao and Tono 2006: 47).

Wyrównanie tekstów w korpusie równoległym nadal pozostaje zadaniem złożonym, napotykanym liczne trudności wynikające z wieloznaczności segmentacji tekstu na zdania (np. różne funkcje kropek, specyfika opisów mowy bezpośredniej, brak wyraźnych granic zdań w zapiskach roboczych) oraz z niezgodności podziału zdań między tekstem oryginalnym a tłumaczeniem, co prowadzi do przesunięć strukturalnych, szczególnie w przypadku tłumaczeń wykonywanych ręcznie (Беляева 2004: 56). Niemniej jednak można wyróżnić trzy podstawowe podejścia do wyrównywania zdań w korpusach równoległych:

- 1) statystyczne,
- 2) lingwistyczne i
- 3) hybrydowe.

Metoda statystyczna opiera się na analizie długości zdań (Brown, Lai and Mercer 1991: 169–176), mierzonej w słowach lub znakach (Gale and Church 1993: 75–102), natomiast

metoda lingwistyczna wykorzystuje informacje morfosyntaktyczne do identyfikacji podobieństw między językami. Chociaż podejście lingwistyczne cechuje się większą precyzją niż statystyczne, jego złożoność i czasochłonność ograniczają stosowanie w dużych korpusach. Z tego względu najczęściej wybiera się metodę hybrydową, która łączy algorytmy probabilistyczne z wiedzą lingwistyczną, zapewniając wyższą skuteczność wyrównania. Teksty wielojęzyczne wyrównane na poziomie zdań stanowią fundament dla automatycznego generowania dwujęzycznych słowników oraz modeli zgodności, które znajdują zastosowanie w systemach przetwarzania języka naturalnego w celu analizy ekwiwalencji i relacji wiarygodności pomiędzy jednostkami leksykalnymi.

Przykładem podejścia do dopasowania ekwiwalentnych zdań jest metoda opracowana przez Gale'a i Churcha, która rozpoczyna się od podziału tekstów w dwóch językach na akapity, w obrębie których przeprowadzane jest automatyczne wiązanie zdań. W każdym akapicie wyróżnia się regiony obejmujące od 0 do 2 zdań w każdym języku, przy założeniu, że pewne typy odpowiadających sobie regionów występują częściej niż inne oraz że dłuższe zdania w języku źródłowym zazwyczaj mają swoje odpowiedniki w postaci dłuższych zdań w języku docelowym. Ową procedurę przeprowadza się przy założeniu istnienia sześciu możliwych typów dopasowania zdań między tekstem źródłowym i jego tłumaczeniem:

- 1) (1:1) jedno zdanie języka L1 odpowiada jednemu zdaniu języka L2,
- 2) (2:1) dwa zdania języka L1 przetłumaczone zostają jednym zdaniem w języku L2,
- 3) (1:2) jedno zdanie języka L1 przetłumaczone zostaje dwoma zdaniami w języku L2,
- 4) (2:2) dwa zdania języka L1 przetłumaczone zostają dwoma zdaniami języka L2, ale ich granice nie odpowiadają sobie,
- 5) (0:1) wstawienie dodatkowego zdania w języku L2,
- 6) (1:0) zdanie w języku L1 nie posiada ekwiwalentu w języku L2 i zostaje pominięte w przekładzie (Lewandowska-Tomaszczyk 2005: 49).

Na podstawie powyższych kryteriów automatyczna procedura wyrównywania wyodrębnia pary tekstowe, odpowiadające zaprezentowanym typom. Dlatego istotnym w procesie podziału tekstu na zdania pozostaje ręczna korekta twórców korpusu, która umożliwia wyznaczenie zakresu ustalenia możliwych dopasowań słów i fraz, a także naniesienie ewentualnych poprawek w tekście. Jak już zostało wspomniane, niektóre korpusy korzystają z algorytmu wykrywania asymetrii tekstowych, które kolejno analizują poszczególne poziomy zestawionych tekstów (rozdziały, akapity, zdania itp.) i w przypadku stwierdzenia dysproporcji sygnalizują wychwycone niedopasowanie. Rozpowszechnioną procedurą w procesie wiązania par przekładowych jest zestawianie odpowiedników na poziomie słowo-słowo, wyrażenie–

słowo oraz słowo–wyrażenie, co umożliwia definiowanie tych jednostek jako funkcjonalne ekwiwalenty fragmentowe. W celu ich automatycznego odnalezienia stosowane są dwie niezależne od siebie procedury:

1) *Strategia totalna*, która polega na traktowaniu tekstu źródłowego jako sekwencyjnego ciągu znaków, a adres każdego słowa jest określony przez odstęp między adresem pierwszego a ostatniego znaku w tekście;

2) *Strategia implikacyjna*, oparta na stopniowym zawężaniu obszaru wyszukiwania w tekście według hierarchii: tom, rozdział, akapit, zdanie, słowo, co pozwala precyzyjnie odnaleźć przypuszczalną jednostkę tłumaczeniową, pod warunkiem zgodności liczby jednostek na każdym poziomie podziału tekstu (Добровольский, Кретов, Шаров 2005: 266).

Korpusy równoległe odgrywają coraz większą rolę w językoznawstwie korpusowym, a jednym z bardziej znanych zbiorów tego typu jest OPUS (Open Source Parallel Corpus), którego zasoby obejmują m.in.: napisy z serwisu opensubtitles.org, przemówienia z posiedzeń Parlamentu Europejskiego, fragmenty Koranu czy pliki lokalizacyjne systemu operacyjnego Linux (Tiedemann 2009: 237–248). Teksty te charakteryzują się dużą różnorodnością gatunkową i tematyczną, przy czym przemówienia Parlamentu Europejskiego dostępne są w ściśle określonej liczbie języków, natomiast zbiór napisów jest systematycznie rozszerzany o nowe wersje językowe. Wszystkie teksty poddawane są automatycznemu przetwarzaniu, obejmującemu proces wyrównywania oraz, w niektórych przypadkach, anotację morfologiczną. Jednak nie są one poddawane ręcznej obróbce. Innym interesującym przykładem wielojęzycznego korpusu równoległego jest korpus Plato, zawierający przekłady *Państwa* Platona na dziesięć języków nowożytnych, w tym angielski, czeski i polski. Jego unikalny charakter wynika z faktu, iż składa się wyłącznie z tłumaczeń, nie zawierając oryginalnego tekstu w języku starogreckim. Zbiór ten został opracowany z dużą precyzją, a jego struktura została szczegółowo opisana w standardowym języku opisu dokumentów SGML. Dzięki temu korpus był wielokrotnie wykorzystywany w badaniach nad algorytmami wiązania ekwiwalentnych jednostek w tekstach równoległych dla różnych języków, w tym również dla języka polskiego. Jednym z najważniejszych korpusów dla języków słowiańskich, takich jak polski, czeski i rosyjski, pozostaje wielojęzyczny zbiór InterCorp, będący integralną częścią NKJCz. Stanowi on bogate i zróżnicowane źródło danych wykorzystywane w badaniach teoretycznych, leksykografii, nauczaniu języków obcych oraz w pracy tłumaczy i naukowców, a jego zasoby są systematycznie powiększane, co umożliwia analizę ekwiwalencji tłumaczeniowej oraz porównawczą analizę struktur językowych. Dzięki dostępności online korpus ten jest cennym narzędziem zarówno dla specjalistów, jak i szerokiej grupy

użytkowników zainteresowanych badaniami językowymi. Umożliwia wyszukiwanie ekwiwalentów dla różnych języków słowiańskich.

Rozwijane są korpusy wielojęzyczne zarówno równoległe, jak i porównywalne. Odgrywają one obecnie kluczową rolę w różnych obszarach językoznawstwa, szczególnie w tłumaczeniu maszynowym, leksykografii oraz różnego rodzaju analizach kontrastywnych i typologicznych. Ich podstawowa funkcja, jaką jest umożliwienie analizy ekwiwalencji między tekstami w różnych językach, znajduje zastosowanie w badaniach lingwistycznych, dydaktyce języków obcych oraz w praktyce tłumaczeniowej. Teksty równoległe umożliwiły identyfikowanie trwałych wyrażen frazeologicznych, idiomów oraz strategii translacyjnych, co nie tylko wzbogaca zasoby leksykograficzne, lecz także wspiera rozwój systemów tłumaczenia maszynowego i glosariuszy wielojęzycznych. Jednym z kluczowych narzędzi w analizie korpusów równoległych są listy konkordancyjne, które umożliwiają obserwację rzeczywistego użycia słownictwa w kontekście, ułatwiając jednocześnie zestawienie odpowiadających sobie jednostek tłumaczeniowych. Jednakże skuteczność tego podejścia zależy od jakości i liczby tekstów włączanych do korpusu, co podkreśla znaczenie reprezentatywności materiału językowego oraz kompetencji twórców i użytkowników takich zasobów. Ponadto, rozwój automatycznych technik rozpoznawania jednostek w tekście pozwala na precyzyjne wyrównywanie jednostek leksykalnych i frazeologicznych w różnych językach, co jest istotne zwłaszcza w kontekście opracowywania wielojęzycznych słowników specjalistycznych, np. frazeologicznych. Równoległe korpusy wielojęzyczne pełnią istotną funkcję w automatycznym przetwarzaniu języka naturalnego, zwłaszcza w statystycznym tłumaczeniu maszynowym, gdzie wykorzystywane są do budowy probabilistycznych tabel tłumaczeń poprzez wyrównanie tekstów zdanie po zdaniu lub fraza po frazie. Ich zastosowanie wykracza jednak poza translatorykę. Stanowią one także cenny materiał do badań kontrastywnych i typologicznych, pozwalając na analizę gramatycznych konstrukcji w różnych językach w oparciu o zdania tłumaczone w sposób zachowujący adekwatność semantyczną i pragmatyczną. Teksty równoległe znajdują również zastosowanie w badaniach nad międzyjęzykowym porównaniem wzorców użycia środków językowych w określonych gatunkach tekstowych. W. Koller podkreśla, że takie badania powinny opierać się na tekstach równoległych, które ujawniają normy klas tekstów w danym języku, co umożliwia poprawne konstruowanie tłumaczeń wykraczających poza sumę poprawnych zdań (Koller 1992: 248).

Pomimo niewątpliwych zalet, przed korpusami wielojęzycznymi stoją poważne wyzwania metodologiczne. Ich fragmentaryczna struktura, wynikająca z faktu, że teksty nie zawsze są dostępne w tych samych językach, prowadzi do nieproporcjonalnej reprezentacji

poszczególnych języków w korpusie. Brak jednolitych oznaczeń morfologicznych i składniowych dodatkowo komplikuje analizę, ponieważ różne języki mogą posiadać odmiennie zakodowane informacje gramatyczne. Konieczność opracowania zunifikowanego schematu anotacji, uwzględniającego specyfikę poszczególnych języków, jest zatem jednym z kluczowych wyzwań w dalszym rozwoju tego działu językoznawstwa. Anotacja automatyczna, choć stanowi nieocenione narzędzie w analizie językowej, może rodzić problemy związane z jakością danych. Błędy w oznaczeniach morfosyntaktycznych i niejednorodność anotacji w różnych językach wpływają na precyzję wyników badań, co w szczególności dotyczy analizy zjawisk gramatycznych w zbiorach tekstowych. Ponadto, obecność kalk, wolnych tłumaczeń oraz błędów wynikających z nieprecyzyjnego wyrównania tekstów może ograniczać przydatność korpusów w badaniach typologicznych i translatologicznych. W przypadku korpusów paralelnych pojawiają się dodatkowe wyzwania, takie jak dobór odpowiednich oryginałów i ich tłumaczeń, co ma istotny wpływ na reprezentatywność i zakres analizy. Często ograniczają się one do jednego lub kilku typów tekstów, takich jak dokumenty prawne, napisy filmowe czy literatura piękna, co wynika przede wszystkim z dostępności danych.

Szczególną trudność stanowią korpusy paralelne języka mówionego, zawierające zarówno oryginalne, jak i tłumaczone wypowiedzi ustne, które w praktyce niemal nie istnieją lub mają ograniczony zasięg gatunkowy, a ich opracowanie jest skomplikowane ze względu na specyfikę mowy spontanicznej i brak standaryzowanych zapisów. Przykładem korpusu tego typu jest Europarl, zawierający transkrypcje wystąpień członków Parlamentu Europejskiego, które, choć bazują na języku mówionym, poddawane są redakcji i formalizacji, co wpływa na ich reprezentatywność (Chlumská 2014: 226). Mimo ograniczeń związanych z dostępnością korpusów paralelnych języka mówionego, teksty beletrystyczne zawarte w korpusach mogą stanowić cenne źródło jednostek charakterystycznych dla języka mówionego (Čermák 2017: 82). Dzięki obecności dialogów oraz stylizacji na mowę codzienną, umożliwiają one analizę wielowyrazowych jednostek leksykalnych, kolokacji oraz idiomatycznych zwrotów typowych dla komunikacji ustnej. Choć nie oddają one w pełni spontanicznego charakteru interakcji werbalnych, ich wykorzystanie w badaniach nad językiem potocznym pozwala na identyfikację tendencji leksykalnych i składniowych, które znajdują odzwierciedlenie w rzeczywistym użyciu języka. Szczególnie istotnymi wielojęzycznymi bazami danych w kontekście badań nad językoznawstwem polskim, czeskim i rosyjskim jest wspomniany korpus InterCorp oraz wielojęzyczny korpus równoległy zawarty w Narodowym Korpusie Języka Rosyjskiego. Oba te zasoby stanowią cenne źródło danych do analizy ekwiwalencji tłumaczeniowej, badań

kontrastywnych oraz modelowania struktur składniowych i leksykalnych w interesujących nas językach.

4.3 Korpus Równoległy InterCorp

4.3.1 Ogólna charakterystyka InterCorp

Jednym z najważniejszych i najbardziej zaawansowanych korpusów uwzględniających interesujące nas języki słowiańskie jest wielojęzyczny korpus równoległy InterCorp. Stanowi on główny projekt o tej samej nazwie, którego głównym celem jest stworzenie obszernego, synchronicznego korpusu równoległego gromadzącego maksymalnie szeroką gamę języków. Owe przedsięwzięcie zostało zrealizowane w 2005 roku na Uniwersytecie Karola w Pradze przy współudziale kadry akademickiej oraz studentów Wydziału Filozoficznego, a także osób związanych z Narodowym Korpusem Języka Czeskiego (Kaczmarska i Rosen 2014: 207). Korpus ten stanowi niezwykle wartościowe źródło danych, wykorzystywane w licznych obszarach badawczych, takich jak językoznawstwo teoretyczne, leksykografia, dydaktyka (ze szczególnym uwzględnieniem nauczania języków obcych) czy w obszarze komputerowego przetwarzania języka naturalnego. Dostęp do InterCorp jest możliwy za pośrednictwem standardowej przeglądarki internetowej poprzez zintegrowany interfejs wyszukiwania CNC, przy czym po dokonaniu rejestracji użytkownicy otrzymują szeroki i bezpłatny dostęp do zasobu. Użytkownicy, oprócz zapytań online kierowanych do internetowego konkordancera, mają możliwość offline'owego dostępu do zestawów równoległych dwujęzycznych konkordancji, które wyodrębniane są z określonych tekstów, zawierają jedynie pary wyrównań 1:1 z losowo przetasowanymi blokami, a dzięki technicznym zabezpieczeniom zgodnym z warunkami licencyjnymi, takimi jak zakaz redystrybucji, umożliwiają dostęp jedynie do cytatów w ograniczonym kontekście, bez naruszania praw autorskich (Čermák i Rosen 2012: 417).

Korpus składa się z dwóch części: *jądra* (czes. *jádro*) oraz *kolekcji* (czes. *kolekce*), które różnią się sposobem pozyskiwania i przetwarzania tekstów. Jądro korpusu InterCorp składa się przede wszystkim z tekstów beletrystycznych, które zostały poddane ręcznej korekcie wyrównania oraz szczegółowej weryfikacji pod kątem literówek, poprawności segmentacji zdań i błędów w procesie wyrównywania. Większość tekstów w zawartych w tym zbiorze przechodzi przez następujące etapy: pozyskiwanie, skanowanie i rozpoznawanie znaków (OCR), korektę, segmentację (wykrywanie granic zdań), wyrównywanie zdań, ponowną korektę oraz sprawdzanie segmentacji i wyrównania, a także oznaczanie morfosyntaktyczne

(Rosen 2016: 35). Natomiast wspomniane kolekcje tworzą teksty pozyskane w wielu językach, które zostały przetworzone i wyrównane automatycznie, co wiąże się z wysokim prawdopodobieństwem wystąpienia błędów. W rezultacie w konkordancjach mogą pojawiać się zdania, które wzajemnie sobie nie odpowiadają. Ponadto należy podkreślić, iż kolekcje nie zawierają wszystkich tekstów z oryginalnych źródeł, szczególnie tych, które nie posiadają czeskiego odpowiednika. Do najnowszej wersji InterCorp v16ud z 2024 roku należą:

- 1) teksty prawne Unii Europejskiej z korpusu *Acquis Communautaire*,
- 2) zapisy posiedzeń Parlamentu Europejskiego z 2007-2011 z korpusu *Europarl*,
- 3) artykuły publicystyczne i informacje z serwisu internetowego *VoxEurop*,
- 4) napisy filmowe z bazy danych *OpenSubtitles*,
- 5) teksty polityczne z bazy *Project Syndicate*,
- 6) przekłady *Biblii*.

Każdy tekst obcojęzyczny w korpusie posiada swoją wersję w języku czeskim, który pełni rolę tzw. *pivot*, czyli tekstu osiowego, będącego oryginałem lub tłumaczeniem, wyrównanym z jedną lub kilkoma wersjami obcojęzycznymi. Należy podkreślić, iż nowe wersje korpusu są udostępniane raz w roku, co pozwala na systematyczny rozwój jego zasobów, a każda kolejna wersja jest większa od poprzedniej. Ponadto zwiększa się praktycznie liczba języków oraz ulega poprawie jakości anotacji. W najnowszej wersji, oznaczonej jako 16ud, korpus zawiera około 4,8 miliarda słów w 61 językach, w tym 516 milionów słów w języku czeskim.

Z perspektywy badań nad wielowyrazowymi jednostkami języka mówionego, takimi jak fatyzmy i emotywy w języku polskim, czeskim oraz rosyjskim, istotne jest, że przytoczone kolekcje nie zawierają tekstów we wszystkich tych językach, co przekłada się na zróżnicowaną przydatność omawianego korpusu w analizie lingwistycznej. Wartość InterCorp w badaniach leksykograficznych i przekładowych zależy natomiast w dużej mierze od rozmiaru zbiorów tekstów dostępnych dla poszczególnych języków. Polsko-czeska wersja korpusu InterCorp pozostaje jednym z największych równoległych zasobów tekstowych, rejestrując ponad 227 mln wyrazów, a wielkość poszczególnych kolekcji przedstawia się następująco: Jądro (ponad 28 mln wyrazów); *Acquis Communautaire* (blisko 19,5 mln wyrazów); *Europarl* (prawie 13 mln); *VoxEurop* (ponad 2,3 mln); *OpenSubtitles* (164 mln wyrazów); Przekłady *Biblii* (0,5 mln wyrazów). Jak można zauważyć, najnowsza wersja (jak i poprzednie) nie zawiera wyłącznie kolekcji tekstów *Project Syndicate*, czyli publicystyki w języku polskim. Z kolei rosyjska wersja korpusu InterCorp wyróżnia się znacznie mniejszą liczbą tekstów oraz brakiem

niektórych kolekcji, a jej zasoby w porównaniu z odpowiednimi wersjami korpusu dla języka czeskiego i polskiego przedstawiono w poniższej tabeli:

Język	Jądro	Acquis	Europarl	VoxEurop	Subtitles	Syndicate	Biblia	Ogółem
CZ	119 681	19 188	12 918	2 313	232 969	4 718	562	397 352
PL	28 093	19 482	12 662	2 367	164 059	-	576	227 242
RUS	12 418	-	-	-	104 831	4 312	565	122 130

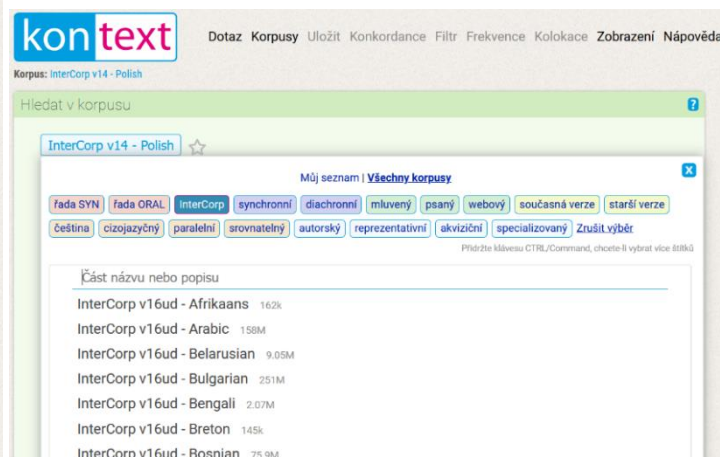
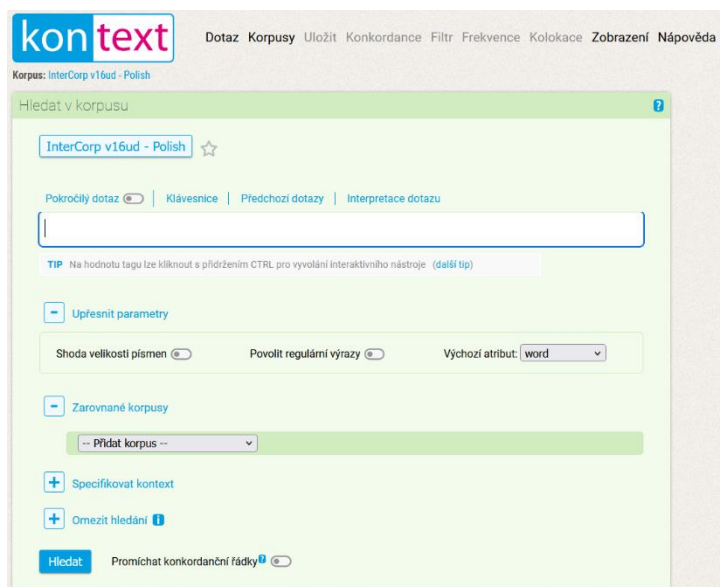
Na podstawie przedstawionej tabeli można dojść do wniosku, iż korpus rosyjski, w przeciwieństwie do zasobów tekstowych korpusu polskiego, nie zawiera niektórych kolekcji, a liczba tekstów w nim zgromadzonych jest również znacząco mniejsza. Istotne różnice w objętości przedstawionych kolekcji oraz ograniczona dostępność kolekcji równoległych nie powinny stanowić przeszkody dla badaczy podejmujących analizę lingwistyczną z wykorzystaniem zasobów InterCorp. Warto zwrócić uwagę na fakt, że wszystkie omawiane tu korpusy słowiańskie zawierają kolekcję OpenSubtitles, która, choć nie może być uznana za wiarygodny zapis mowy ustnej (potocznej), stanowi materiał tekstowy częściowo odzwierciedlający spontaniczne wypowiedzi w nieformalnych kontekstach (Zięba 2022: 203). Dostępne kolekcje mogą więc stanowić bazę do identyfikacji potencjalnych jednostek fatycznych i emotywnych oraz ich tłumaczeń.

4.3.2 Wyszukiwanie w InterCorp

Jednym z kluczowych narzędzi zintegrowanych z każdym zbiorem tekstowym czeskiego korpusu (w tym InterCorp) jest wyszukiwarka KonText. Stanowi ona zaawansowany instrument do analizy i eksploracji danych językowych, opracowana została w ramach projektu LINDAT/CLARIAH-CZ. Aplikacja webowa pozwala użytkownikom na formułowanie zarówno nieskomplikowanych, jak i bardziej zaawansowanych zapytań, a wyniki wyszukiwania przedstawiane są w postaci listy konkordancji. Jedną z istotnych funkcji KonText jest możliwość obliczania rozkładów częstościowych, co umożliwia statystyczną analizę danych językowych, oraz identyfikacji kolokacji za pomocą miar asocjacyjnych, przydatnej w badaniach nad kolokacjami czy związkami frazeologicznymi. Aplikacja oferuje również różnorodne opcje filtrowania i sortowania wyników, umożliwiając dostosowanie prezentacji danych do indywidualnych potrzeb badawczych, a także wspiera pracę z różnymi

typami korpusów, w tym z korpusami równoległymi, co jest szczególnie użyteczne w badaniach porównawczych.

Proces przeszukiwania korpusu InterCorp cechuje się prostotą interfejsu użytkownika i umożliwia zarówno analizę wskazanego zapytania w wybranym zbiorze jednojęzycznym (dla konkretnego języka), jak i eksplorację dwóch korpusów równoległych, co pozwala na uzyskanie potencjalnych par tłumaczeniowych do dalszej weryfikacji. Pierwszym etapem pracy z korpusem jest wybór z menu początkowego (które zawiera wszystkie zbiory tekstowe wchodzące w skład NKJCz) etykiety InterCorp oraz wskazanie konkretnego języka do analizy. Dzięki możliwości przeszukiwania wybranej jednostki lub reproduktu wielowyrazowego w określonym zbiorze tekstowym dla danego języka, korpus ten można traktować jako jednojęzyczny, który umożliwia m.in. oszacowanie frekwencji danego wyrażenia, analizę jego semantyki oraz badanie łączliwości z innymi leksemami. W celu uzyskania konkordancji równoległych dla interesujących nas języków, konieczne jest dodanie kolejnego języka w sekcji *Zarovnané korpusy*, co pozwala na zestawienie tekstów w dwóch lub więcej językach i przeprowadzenie szczegółowej analizy porównawczej:



W repertuarze pozostałych dostępnych funkcji tego korpusu znajdują się zaawansowane parametry, umożliwiające precyzyjne dostosowanie procesu wyszukiwania:

1. *Pokročilý dotaz* - umożliwia użytkownikowi formułowanie złożonych zapytań przy użyciu języka CQL (Corpus Query Language), oferując automatyczną kontrolę i

podświetlanie składni wprowadzanych zapytań oraz informując o ewentualnych błędach składniowych przed ich przetworzeniem.

2. *Klávesnice* - interaktywna międzynarodowa klawiatura, umożliwiająca wprowadzanie znaków specjalnych, co jest szczególnie przydatne podczas wyszukiwania w tekstach obcojęzycznych lub korzystania ze specjalnych symboli w języku CQL.
3. *Předchozí dotazy* - sekcja umożliwia przegląd i filtrowanie ostatnio zadanych zapytań według używanego korpusu lub typu, z opcją wyświetlania zarchiwizowanych zapytań, które można ponownie wprowadzić do formularza w celu natychmiastowego przetworzenia lub dalszej modyfikacji.
4. *Interpretace dotazu* - pozwala użytkownikowi sprawdzić, w jaki sposób jego zapytanie zostanie przetworzone (czyli przetłumaczone na język CQL) oraz ocenić, czy interpretacja ta odpowiada zamierzonemu celowi, np. po wpisaniu zapytania *Co tím myslíš* urządzenie przedkłada następującą interpretację:

word="(?)co" | word="(?)tím" | word="(?)myslíš"

5. *Upřesnit parametry* – funkcja ta umożliwia określenie dodatkowych parametrów zapytania, takie jak *Shoda velikosti písmen* (uwzględnienie wielkości liter), *Povolit regulární výrazy* (możliwość korzystania z wyrażeń regularnych) czy *Výchozí atribut* (wybór atrybutu pozycyjnego, np. word, lemma, case itp.). W przypadku wyszukiwania zapytania w wybranym korpusie równoległym z innego języka pojawia się dodatkowa opcja *Aplikace dotazu* (wybór czy wyniki powinny zawierać odpowiadające tłumaczenia) oraz *Včetně nezarovnaných* (przełącznik ten kontroluje, czy system ma również przeszukiwać fragmenty tekstów, które nie są idealnie zrównane między językami w korpusie).
6. *Specifikovat kontext* – możliwość dodatkowego sprecyzowania zapytania w zależności od kontekstu (tekstowego otoczenia), w którym występuje poszukiwane słowo lub fraza. W ramach tej funkcji użytkownik może określić: *rozsah kontextu* (zakres kontekstu), czyli odległość, w jakiej mają znajdować się wyrazy względem kluczowego słowa oznaczonego jako KWIC; *seznam lemmat oddělených mezerami* (listę lematów oddzielonych spacjami), tj. zestaw lematów uwzględnianych w analizie; a także wybrać kryterium uwzględniania lematów, które pozwala na uwzględnienie bądź odrzucenie wszystkich wpisanych jednostek, np. *musí být přítomny všechny položky*.
7. *Omezit hledání* – pozwala na zawężenie zakresu zapytania użytkownika do określonych zbiorów tekstowych (np. literatury beletrystycznej lub napisów filmowych). Proces ten

może być dodatkowo precyzowany poprzez wybór m.in.: konkretnego autora, tytułu lub tekstu oryginalnego.

8. *Promíchat konkordanční řádky* – Autorzy korpusu zalecają pozostawić wskazaną funkcję włączoną, gdyż zapewnia losową kolejność wyświetlania konkordancji, co jest istotne dla poprawnej interpretacji wyników. Jednak w przypadku wyszukiwań obejmujących bardzo duże korpusy (np. miliard tokenów) można ją tymczasowo wyłączyć, aby uniknąć znacznego wydłużenia czasu oczekiwania na rezultaty.

W celu uzyskania paralelnych konkordancji dla wybranych języków, najpierw należy wybrać język wyjściowy, którym w naszym przypadku będzie najnowsza wersja korpusu polskiego, sformułować zapytanie (np. *Ani mi się śni*), oraz wybrać czeską wersję korpusu w celu identyfikacji potencjalnych ekwiwalentów. Należy podkreślić, że narzędzie automatycznie oblicza najczęstsze odpowiedniki wskazanego zapytania jednowyrazowego za pomocą aplikacji Treq, które są wyświetlane po lewej stronie listy konkordancji. Poniżej przedstawiono przykłady równoległych konkordancji polsko-czeskich uzyskanych dla wskazanego zapytania, które ilustrują potencjalne ekwiwalenty językowe w obu korpusach oraz ich konteksty użycia:

konTEXT			
Dotaz Korpusy Uložit Konkordance Filtr Frekvence Kolokace Zobrazení Návoděda			
Korpus: InterCorp v16ud - Polish Dotaz: ani, mi, się, śni (134 výskytů) ~ Podrobnosti			
Vyskytl: 134 l.p.m.: 0,45 (vřazeno k celému korpusu) ARF: 73,39 Výsledek je setříděn			
Výběr řádků: základní			
InterCorp v16ud - Polish		InterCorp v16ud - Czech	
☐ bouskova-rok_kohout	- Ani mi się śni ! - powiedzialem i pobiegam po ciepłe rajtuzy i ciepłe skarpety , zeby po drodze nie zmarznęć .	bouskova-rok_kohout	Ani mě nenapadne ! Řekla jsem a běžela pro teplé punčocháče a teplé ponožky , aby mi cestou nebylo zima .
☐ caroll-alenka_zrcadle	* Ani mi się śni ! - rzekła już zniecierpliwiona Alicja .	caroll-alenka_zrcadle	„ A nekoupim ! “ utrhla se na něj Alenka , „ vždyť já do tohohle vlaku vůbec nepatřím - byla jsem v lese - kdybych tak mohla zpátky . “
☐ Clavel-TchajPan	- Ani mi się śni !	Clavel-TchajPan	* Něco takového neudělám . *
☐ Frankova-DenikA_Franko	Ani mi się śni dawać się obrzucać tymi obelgami , ja im pokażę , że Anne Frank nie jest niedzisiejsza , podniosą jeszcze głowę i prędko zamkną głęby , gdy im uświadomię , że nie od mojego wychowania muszą zacząć , ale od wychowania samych siebie .	Frankova-DenikA_Franko	Ani mě nenapadne , abych si nechala libit takové urážky . Já jim ukážu , že Anne Franková není včerejší ! Budou se ještě divit a zavřou ty svoje klapáčky , až jim objasním , že musí začít ne s mou výchovou , ale nejdřív se svou vlastní .
☐ Grass-Sire_gole	- Ani mi się śni .	Grass-Sire_gole	„ Ani ve smu by mě nenapadlo . “
☐ Halley-Letiste	- Ani mi się śni ! - odkrzyknęł .	Halley-Letiste	* Konina , Bakersfelde !
☐ Hasek-OsudyDobrohoVSV	- Nie będę - powtarzał uparcie - nic dla was nie uczynię , ani mi się śni , bo jesteście niepoprawni nikczemnicy .	Hasek-OsudyDobrohoVSV	* A nebudu , * opakoval umíněné , * nic pro vás neudělám , ani mne nenapadne , poněvadž jste nenapravitelní ničemové .
☐ Heller-Hlava_XXXI	- Ale ani mi się śni .	Heller-Hlava_XXXI	* Ale ani mě nenapadne , abych něco takového udělal . *
☐ Koontz-Podivne_cesty	- Ani mi się śni .	Koontz-Podivne_cesty	* Leda bych se zbližmila . *
☐ meyerova-stmivani	- Ani mi się śni - przerwał mu Emmett .	meyerova-stmivani	* Ani nápad , * vpadl do toho Emmett .
☐ Parál-SoukromiVichrice	... a Bohunka powiada na to : ani mi się śni , sami smarkacze , dla mnie w całej fabryce jest tylko jeden prawdziwy męszczyna , Standa Kocián ... - wysłał się kumpel Adzik , aż ochrypl .	Parál-SoukromiVichrice	* ... a Bohunka na to prej Kdepak , to jsou sami strabaři , jedinej opravdovskej chlap ve fabrice je u mě Standa . Kocián ... , namáhal se kámoš Áda , až chraptěl .
☐ remarque-tr_kamaradi	- Ale gdzie tam , ani mi się śni chorować .	remarque-tr_kamaradi	* Ale kde , ani trochu .
☐ rowlingova-hpot_pohar	- Ani mi się śni ! - krzyknął Harry , ale tylko cztery wielkie bąble wyleciały mu z ust .	rowlingova-hpot_pohar	* A to zas ne ! * vybuchl Harry hněvivě - z úst mu však vyklouzly jen dvě velké bubliny .

Możliwe jest również przeprowadzenie szczegółowej analizy zdań, w których reprodut *Ani mi się śni* został przetłumaczony w określony sposób, np. jako *Ani mě nenapadne*. W tym celu należy wprowadzić w obu językach odpowiednie jednostki wielowyrazowe, co pozwoli na uzyskanie wyników zawierających wskazane pary tłumaczeniowe:

KOTLEX

Korpus: InterCorp v16ud - Czech | Dotaz: co, tím, myslíš (14 821 výskytů) ~ Pr

Výskytů: 14 821 | I.p.m.: 28,68 (vztaženo k celému korpusu) | ARF: 6 900,5

Výběr řádků: základní

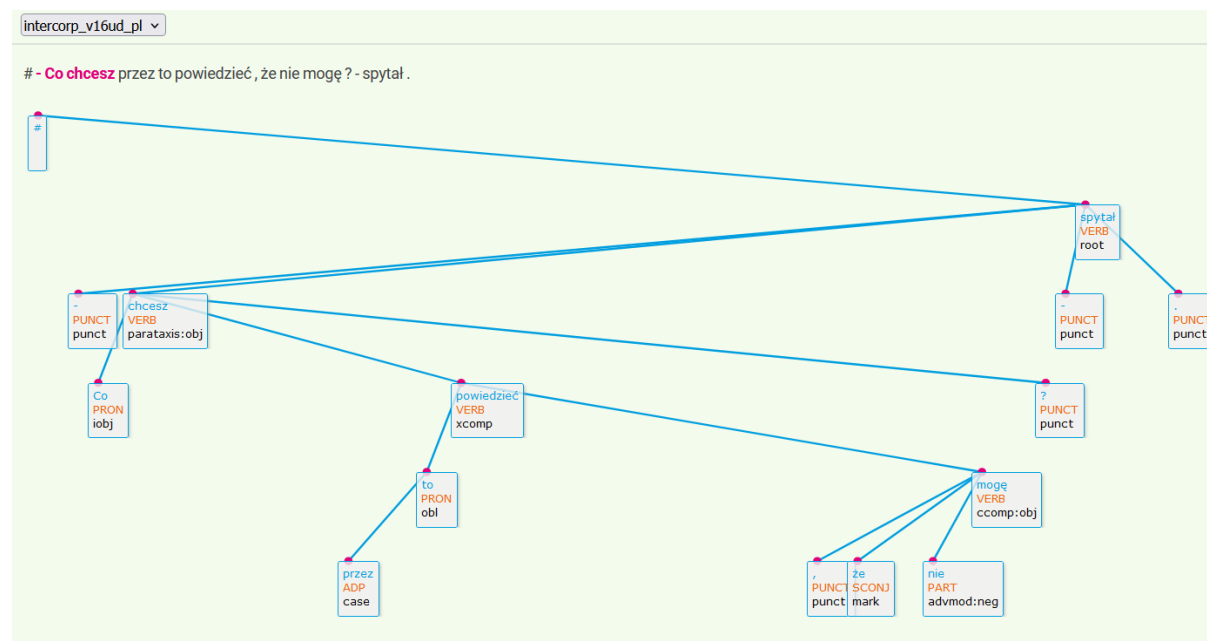
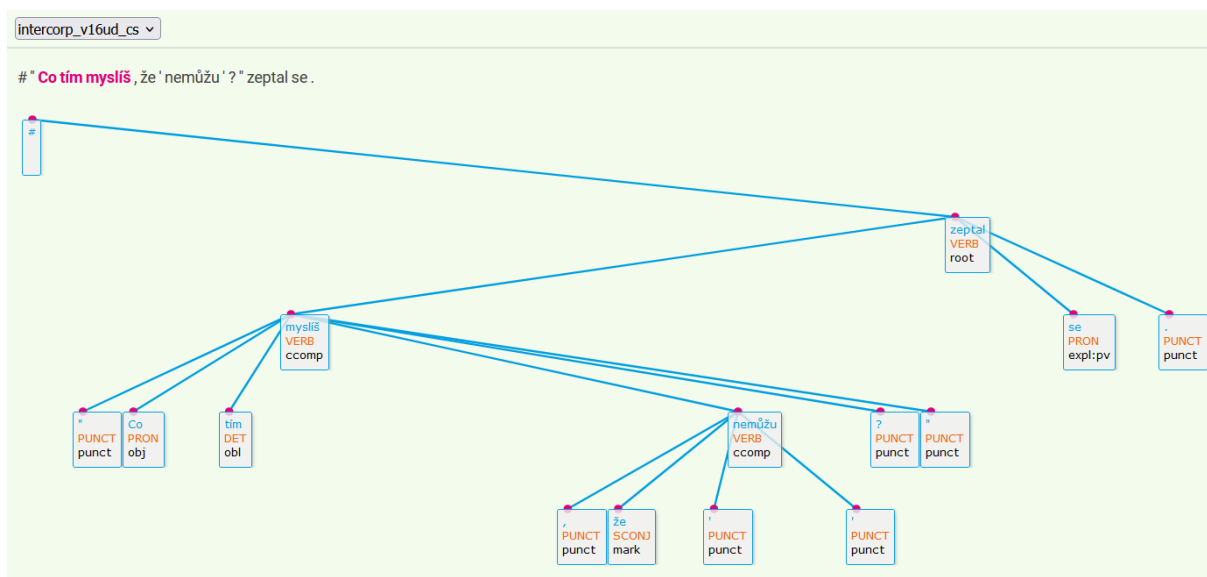
☐	abele-vitr
☐	adams-prevazne_neskon
☐	adams-prevazne_neskon
☐	adams-prevazne_neskon
☐	adams-prevazne_neskon
☐	adams-prevazne_neskon
☐	adams-prevazne_neskon
☐	adams-sbohnm_a_6lk
☐	edichie-amerikana
☐	edichie-amerikana
☐	edichie-amerikana
☐	edichie-amerikana
☐	edler-slozka4
☐	edler-slozka4
☐	edler-slozka4
☐	edler-vytkaz_v_jahvi
☐	edler-zabijaci
☐	Ajar-Zivot_pred_sebou
☐	Ajmatov-Poproviste
☐	Ajmatov-Poproviste

doc.id	abele-vitr	doc.tag_model	czech-fictree-ud-2.12-230717
text.lang	cs	text.pubyear	==NONE==
text.version	0	text.pubmonth	==NONE==
text.pubDateYear	2009	text.pubDateMonth	
text.id	cs:abele-vitr:0	text.author	Abele, Inga
text.title	Vitr proměnlivých směrů	text.group	Core
text.publisher	Kniha Zlín	text.pubplace	Zlín
text.origyear		text.isbn	978-80-87162-56-9
text.txttype	fiction	text.comment	
text.original	No	text.srclang	lv
text.translator	Matoušková, Lenka	text.transsex	F
text.authsex	F	text.transcomment	
text.collectionauthor		text.collectiontitle	
text.volume		text.pages	
text.lang_var	cs-CZ	text.wordcount	57005
text.lexDivWord	645,88	text.lexDivLemma	502,86
text.subRatioAvg	1.50	text.maxTreeDepthAvg	0.57
text.sLengthAvg	10.56	text.mdd	2.39
text.maxNPLengthAvg	3.59	text.maxNPDepthAvg	1.30
p.id	cs:abele-vitr:0:1277	s.id	cs:abele-vitr:0:1277:1
s.maxNPDepth	0	s.subRatio	1.00
s.sLength	3	s.maxNPLength	1
s.mdd	1.50	s.maxTreeDepth	0
hi.rend		lb.id	
div.id		div.type	

Z kolei w przypadku anotacji wewnętrznej należy wziąć pod uwagę fakt, że system znakowania tekstów przeszedł istotne zmiany w kolejnych wersjach korpusu, przy czym w najnowszej wersji 16vud 47 spośród 61 uwzględnionych języków zostało oznakowanych zarówno morfologicznie, jak i syntaktycznie. Do tej pory każdy z języków charakteryzował się znacznym zróżnicowaniem zestawów tagów oraz niespójnością zasad tokenizacji i lematyzacji, co wynikało z pragmatycznego podejścia polegającego na wykorzystaniu dostępnych narzędzi, takich jak tokenizatory, tagery i lematyzatory opracowane przez ekspertów, a także modeli anotacji tworzonych i trenowanych w ramach innych projektów (Rosen 2016: 36). Obecnie wszystkie języki wykorzystują spójny system anotacji zgodny ze standardem Universal Dependencies, którego szczegółowe zasady znakowania zostały opisane na stronie: <https://universaldependencies.org/guidelines.html>. [data dostępu 25.12.2024].

Dzięki zastosowaniu omawianego narzędzia użytkownik uzyskuje szczegółowe informacje dotyczące takich aspektów jak m.in.: części mowy, kategorie morfologiczne oraz funkcje syntaktyczne analizowanego zapytania. Należy również podkreślić, iż najnowsza wersja korpusu jako jedyna do tej pory obejmuje metryki złożoności składniowej (*sintaktická komplexita*) i różnorodności leksykalnej (*lexikální diverzita*). Pierwsza z nich odnosi się do poziomu strukturalnego skomplikowania zdania lub tekstu, który mierzy stopień organizacji gramatycznej oraz relacji między poszczególnymi składnikami syntaktycznymi, takimi jak frazy, zdania podrzędne i różnorodne struktury składniowe, a jej analiza często obejmuje badanie złożoności hierarchicznej, liczby jednostek syntaktycznych oraz ich interakcji w kontekście lingwistycznym. Druga z kolei odnosi się do stopnia zróżnicowania zasobów leksykalnych w danym tekście i jest mierzona poprzez analizę liczby unikalnych jednostek leksykalnych (typów) w relacji do całkowitej liczby wyrazów (tokenów), co pozwala ocenić

bogactwo słownictwa, stopień powtarzalności oraz kreatywność językową w kontekście podejmowanej analizy. Poniżej zaprezentowano próbkę anotacji w formie drzewa zależności, uwzględniając system znakowania dla zapytania *Co tím myslíš?* w czeskiej wersji korpusu oraz jego polskiego odpowiednika tekstowego, również poddanego anotacji:



4.3.4 Zalety i ograniczenia korpusu InterCorp

InterCorp oferuje unikalne możliwości analizy tłumaczeń dzięki dostosowaniu tekstów do *tertium comparationis*, czyli wspólnego punktu odniesienia umożliwiającego porównanie oryginału z jego odpowiednikiem w języku docelowym. Należy jednak zaznaczyć, że zdecydowana większość tłumaczeń dostępnych w InterCorp dotyczy relacji angielski - czeski i angielski - inne języki, co prowadzi do podwójnej konceptualizacji tej samej sytuacji przez tłumaczy różnych języków. W takim układzie oba tłumaczenia funkcjonują równolegle jako niezależne produkty językowe, a nie warianty wtórne względem jednego oryginału, co komplikuje ocenę ekwiwalencji. Różnorodność tłumaczeń wynikająca z podwójnej konceptualizacji powoduje, że użytkownik korpusu nie uzyskuje gotowych odpowiedników, lecz otrzymuje obszerny materiał językowy do samodzielnych badań. InterCorp stwarza możliwość szczegółowej analizy ekwiwalencji tłumaczeniowej oraz różnic systemowych między językami, zwłaszcza w zakresie zasobu leksykalnego, związków frazeologicznych czy struktur składniowych (Charciarek 2019: 196–198). Z punktu widzenia procesu tłumaczeniowego istotne jest, że InterCorp precyzyjnie wskazuje kierunek przekładu, czyli identyfikuje język źródłowy, co stanowi kluczowy czynnik techniczny wpływający na jakość tłumaczenia. Ponadto możliwość przeprowadzenia tłumaczenia odwrotnego pozwala na dodatkową weryfikację przyjętych rozwiązań translatorskich, umożliwiając wykrycie potencjalnych błędów, w tym wynikających z interferencji językowej.

Można stwierdzić, że obszerne zasoby korpusu InterCorp w pewnym stopniu odzwierciedlają cechy autentycznego języka mówionego. Dzięki temu korpus ten może stanowić wartościowe narzędzie w badaniach nad wielowrazowymi reproduktami charakterystycznymi dla tego rejestru języka. Należy jednak mieć na uwadze, że w rzeczywistości nie są to teksty w pełni autentyczne, lecz stylizowane, co widoczne jest chociażby na przykładzie dialogów postaci w utworach prozatorskich. Podobne ograniczenia dotyczą także napisów filmowych, które stanowią zapis wypowiedzi w codziennych sytuacjach komunikacyjnych. Szczególnie amatorskie napisy filmowe charakteryzują się licznymi mankamentami, wynikającymi z faktu, że w większości stanowią tłumaczenia z innych języków, przede wszystkim z języka angielskiego. W konsekwencji badania oparte na takich danych często sprowadzają się do porównywania dwóch tłumaczeń, co może wpływać na dokładność czy reprezentatywność wyników.

Nie należy zapominać o szeregu ograniczeń analizowanego korpusu, które znacząco wpływają na jego funkcjonalność i użyteczność badawczą. Jednym z podstawowych

problemów jest ograniczona różnorodność tekstów, co wynika nierzadko z preferencji koordynatorów odpowiedzialnych za dane języki oraz braku dostępności odpowiednich tłumaczeń. Taka sytuacja połączona z różnicami w objętości danych w korpusie, determinowanymi zarówno dostępnością tekstów, jak i priorytetami badawczymi, prowadzi do powstania materiału, który bywa jednocześnie niespójny i nie w pełni reprezentatywny. Istotnym ograniczeniem technicznym jest z kolei brak możliwości obsługi wielu tłumaczeń w ramach jednego języka, co stanowi przeszkodę dla użytkowników zainteresowanych badaniami nad różnorodnymi strategiami tłumaczeniowymi. W tym kontekście warto także zwrócić uwagę na brak funkcji biKWIC⁸, która mogłaby podkreślać równoważniki słów kluczowych w innych językach, umożliwiając bardziej precyzyjne analizy. Dodatkowo, chociaż dane korpusu zawierają informacje o wyrównaniach (np. 1:1, 2:1, 1:2), brakuje zarówno oznaczeń potwierdzających ich weryfikację przez człowieka, jak i wskaźników pewności generowanych automatycznie, co znacząco ogranicza ich wiarygodność podczas interakcji użytkownika z korpusem (Hebal-Jezierska, Kaczmarek, Rosen 2016: 48).

Kolejną kwestią są trudności, jakie napotykać użytkownicy, zwłaszcza początkujący, którzy mogliby skorzystać z bardziej rozwiniętej pomocy kontekstowej i mechanizmów wspierających analizę. Obecna infrastruktura nie oferuje funkcji takich jak wykresy obrazujące strukturę wybranego korpusu, szczegółowe informacje o typach tekstów ani automatyczne generowanie zapytań w języku CQL. Brak także zaawansowanych opcji analitycznych, takich jak porównania typów tekstów, języków czy profili kolokacyjnych, co ogranicza potencjał badawczy narzędzia. Warto dodać, że uprzywilejowanie literatury pięknej jako źródła tekstów, choć zapewnia bogactwo językowe, zmniejsza reprezentatywność materiału w kontekście analizy różnorodnych gatunków dyskursów. Analiza języka mówionego na podstawie zbiorów pisanych tekstów równoległych dostępnych w tym korpusie stawia przed badaczami istotne wyzwania metodologiczne i interpretacyjne, wynikające z fundamentalnych różnic między formami pisanymi a mówionymi. Niemniej jednak przy odpowiednim podejściu badawczym oraz zastosowaniu zaawansowanych narzędzi analitycznych, prowadzenie tego typu badań jest możliwe i może dostarczyć adekwatnych par przekładowych. Częstym problemem pozostaje również przenoszenie nawyków wypracowanych podczas pracy z korpusami jednojęzycznymi na korpus równoległy, co prowadzi do nierealistycznych oczekiwań wobec dostępnego

⁸ Funkcja **biKWIC** (bilingual Key Word in Context) jest narzędziem stosowanym w korpusach równoległych, które umożliwia jednoczesne wyświetlanie kontekstów dla kluczowych słów w dwóch językach. Dzięki temu użytkownicy mogą łatwiej identyfikować równoważniki słów kluczowych w języku docelowym, co pozwala na bardziej precyzyjną analizę porównawczą tekstów.

oprogramowania. Niedostateczne doświadczenie użytkownika może skutkować niską liczbą wystąpień zapytania, ograniczeniem zakresu tematów badawczych oraz uzyskaniem wyników o ograniczonej wartości naukowej. Dodatkowym wyzwaniem jest nieuwzględnianie wpływu kierunku tłumaczenia, brak zrozumienia różnic w teorii lingwistycznej, a także niewłaściwe zastosowanie metod analitycznych, co może prowadzić do błędnych interpretacji i wniosków. Wykorzystanie tekstów tłumaczonych jako materiału badawczego w korpusie InterCorp wiąże się z istotnym wyzwaniem metodologicznym, wynikającym z faktu, iż tłumaczenia nie są tekstami oryginalnymi, a ich język często odbiega od wzorców użycia w języku docelowym. Użytkownik ma do czynienia ze znaczącymi różnicami w sposobach przekładu tego samego tekstu na różne języki, a także z obecnością tłumaczeń pośrednich z języka trzeciego, co może wpływać na wiarygodność wyników badań (Čermák 2010: 13–14). W związku z tym, mimo że InterCorp jest wartościowym narzędziem badawczym, konieczne są dalsze prace nad jego rozwojem i doskonaleniem, aby lepiej odpowiadał na potrzeby użytkowników i umożliwiał bardziej kompleksową analizę danych.

4.4. Korpus Równoległy Narodowego Korpusu Języka Rosyjskiego

4.4.1 Ogólna charakterystyka

Korpus Równoległy (*Параллельный корпус*) NKJR zawiera dwujęzyczne zbiory tekstowe, w których językiem odniesienia i podstawą wyrównania jest język rosyjski. Ogólną koordynację projektów równoległych korpusów, rozwijanych od 2005 roku, prowadził D. W. Siczinawa, natomiast w realizacji tych prac uczestniczyli programiści, tacy jak T. A. Archangielskij, który opracował m.in. środowisko robocze do wyrównywania tekstów oraz analizator morfologiczny Uniparser, oraz A. P. Pecionyj, odpowiedzialny za implementację wyrównywania zdań w wielojęzycznym korpusie. Większość dostępnych w NKJR korpusów równoległych zawiera zarówno tłumaczenia tekstów obcojęzycznych na język rosyjski, jak i odwrotnie – tekstów rosyjskich na inne języki. Do języków w tej kolekcji należą m.in.: angielski, ormiański, baszkirski, białoruski, bułgarski, hiszpański, włoski, chiński, łotewski, litewski, niemiecki, polski, portugalski, słoweński, ukraiński, fiński, francuski, czeski, czuwaski, szwedzki, estoński oraz japoński. W realizacji projektu uczestniczyło kilka zespołów specjalistów z różnych krajów, a rozwój tego kierunku wspierały również osobne programy grantowe, w ramach których powstawały także niezależne inicjatywy, takie jak polski korpus, finansowany z odrębnych środków. Obecnie ów korpus stanowi jeden z największych projektów NKJR i zawiera ponad 214 milionów słów.

Należy zaznaczyć, iż zasób ten zawiera również wielojęzyczny korpus (*многоязычный корпус*), w którym znajdują się tłumaczenia dzieł literackich z XIX i XX wieku. Teksty te, będące oryginalnie zapisane w różnych językach, zostały przetłumaczone na kilka dziesiątek języków, głównie słowiańskich. Obecnie korpus zawiera 12 tekstów, z których niektóre występują w kilku wersjach, zarejestrowanych w korpusie jako odrębne teksty:

1. *Alicja w Krainie Czarów*; L. Carrolla
2. *Alicja po drugiej stronie lustra* L. Carrolla;
3. *Pies Baskerville'ów* A. Conan Doyle'a;
4. *Kubuś Puchatek* A.A. Milne'a;
5. *Kod Leonarda da Vinci* D. Browna (3 teksty przekładów);
6. *Mały Książe* A. de Saint-Exupéry'ego;
7. *Pinokio* C. Collodiego;
8. *Alchemik* P. Coelho;
9. *Mistrz i Małgorzata* M. Bułhakowa. (2 teksty przekładów).

Każdy z tych tekstów jest przedstawiony w oryginale i w tłumaczeniach (od 6 do 25 tłumaczeń, głównie na języki słowiańskie), a w niektórych przypadkach uwzględniono kilka tłumaczeń na jeden język. U podstaw danego zbioru leżą teksty korpusu ASPAC, w którym wyrównywanie tekstów pierwotnie przeprowadzono na poziomie akapitów, jednak na potrzeby NKJR proces ten został półautomatycznie dostosowany i zastąpiony wyrównywaniem na poziomie pojedynczych zdań. Warto odnotować, że w zasobach tego korpusu równoległego znajdują się również teksty poliwariantywne, czyli takie, które zawierają więcej niż jedno tłumaczenie z języka L1 na język L2, z czego najbardziej rozwijany jest poliwariantywny korpus rosyjsko-francuski, w którym dla wybranych tekstów uwzględniono do czterech różnych wersji przekładu. Teksty poliwariantywne z kolei są także uzupełnieniem korpusu wielojęzycznego, a plany rozwoju zakładają zastosowanie analogicznego podejścia do innych par językowych pochodzących z najbardziej obszernych korpusów, takich jak angielski i niemiecki.

Zawarte zbiory równoległe zawierają obecnie zarówno teksty literackie, jak i nieliterackie, opierając się na tej samej klasyfikacji, co w głównym korpusie języka rosyjskiego. Zawiera ona takie gatunki tekstów, jak publicystyka, teksty techniczno-produkcyjne, naukowo-dydaktyczne, religijne czy prawne. Obecnie rozszerza się dane zasoby o materiały niebeletrystyczne, co pozwala na wierniejsze odwzorowanie współczesnej leksyki i terminologii w porównaniu z tekstami literackimi, dzięki mniejszej swobodzie translatorskiej. Szczególną wartość mają jednak teksty ukazujące regionalne odmiany języków literackich,

takie jak wschodni i zachodni wariant języka ukraińskiego. W ostatnich latach intensywnie rozwijane korpusy językowe są systematycznie wzbogacane o literaturę oraz teksty publicystyczne, reprezentujące zarówno geograficzną (np. proza kastylijska), jak i kulturową różnorodność języków, zawierając zarówno ich historyczne, jak i współczesne warianty. Wśród trudności należy wymienić niejednoznaczność między językiem oryginału a językiem przekładu (np. w dwujęzycznych instrukcjach dotyczących leków, tekstach internetowych mediów itp., gdzie kierunek tłumaczenia można ustalić jedynie poprzez mniej lub bardziej prawdopodobne założenia). Dodatkowo pojawiają się wątpliwości co do lingwistycznej poprawności tekstów wynikających z ręcznej edycji tłumaczeń maszynowych. Można wysnuć wniosek, iż obecny korpus, podobnie jak inne korpusy równoległe, cechuje się ograniczoną reprezentatywnością, wynikającą z opierania się na tekstach już przetłumaczonych, co prowadzi do zniekształcenia proporcji pomiędzy różnymi zbiorami tekstowymi. Poniżej przedstawiono próbkę danych obejmującą zasoby tekstowe, liczbę użyć słów oraz procentowy udział w całym korpusie dla najbardziej reprezentatywnych języków (angielski, niemiecki, szwedzki), dla pozostałych języków wschodniosłowiańskich (białoruski, ukraiński) oraz dla interesujących nas zbiorów w języku polskim i czeskim:

Korpus	Liczba tekstów	Liczba użyć słów	Procentowy udział w całym korpusie
Polski	54	6 355 629	0.3%
Czeski	556	4 387 470	0.2%
Angielski	1 556	51 982 439	2.3%
Niemiecki	299	32 276 755	1.5%
Białoruski	312	10 916 697	0.5%
Ukraiński	865	9 383 774	0.4%
Szwedzki	787	16 520 152	0.7%

Niewątpliwie zauważalna jest znacząca dysproporcja między poszczególnymi zbiorami dostępnymi w korpusie równoległym, przy czym zastanawiający jest fakt, że czeski korpus, mimo posiadania większej liczby tekstów niż polski, charakteryzuje się niższą liczbą użyć słów. Prace nad korpusem polskim były prowadzone w ramach wspólnego projektu z Uniwersytetem Warszawskim pod kierownictwem M. Łazińskiego, i wbrew pozorom stanowi jedną z największych par językowych, liczącą łącznie ponad 6,3 miliona wystąpień słów. Należy podkreślić, że korpus dobrze odzwierciedla polską prozę literacką różnych okresów, od twórczości Sienkiewicza i Prusa po dzieła Chmielewskiej i Sapkowskiego, a ponadto zawiera również artykuły prasowe oraz teksty prawne. Rosyjsko-polski korpus równoległy nie jest jednak jeszcze w pełni reprezentatywny, ale zawiera m.in. materiał nietypowy dla korpusów

równoległych, taki jak proza rosyjskiego srebrnego wieku (np. twórczość Briusowa i Sołoguba) we współczesnych polskich przekładach (Сичинава 2015: 202–203). Z kolei język czeski przez długi czas pozostawał istotną luką w NKJR, co stanowiło wyzwanie szczególnie w kontekście obecności wielu innych języków słowiańskich uwzględnionych w ramach korpusów dwujęzycznych oraz intensywnego rozwoju lingwistyki korpusowej w Czechach, zwłaszcza w zakresie korpusów równoległych. Dopiero w 2018 roku, dzięki inicjatywie T. A. Małyszewy rozpoczęto prace nad korpusem czeskim, który oparto na danych pochodzących z równoległego korpusu ASPAC autorstwa A. Barentsena. Jednocześnie nawiązano współpracę z korpusem InterCorp, rozwijanym w ramach Narodowego Korpusu Języka Czeskiego, co pozwoliło na wzajemną wymianę tekstów i wzbogacenie zasobów językowych (Сичинава 2019: 47). W procesie morfologicznego znakowania tekstów czeskojęzycznych wykorzystano zaawansowany analizator opracowany przez firmę Yandex. Aktualnie zbiory zawierają zarówno współczesne teksty prasowe, jak i klasyczne dzieła literackie, w tym czteroczęściową powieść *Nieznośna lekkość bytu* Milana Kundery oraz twórczość autorów takich jak Puszkina, Czechow, Bułhakow, Pietrow czy Nabokov.

4.4.2 Wyszukiwanie w korpusie równoległym i wielojęzycznym

Struktura wyszukiwarki w korpusie równoległym, choć zbliżona do mechanizmów stosowanych w innych korpusach i wyposażona w niektóre analogiczne funkcje, charakteryzuje się istotnymi różnicami wynikającymi z jej specyfiki badawczej w zakresie analizy dwóch języków, które zostaną w tym miejscu szczegółowo omówione. Pierwszą z funkcji jest znana z poprzedniego rozdziału funkcja *Поиск точных форм*, która pozwala na wyszukiwanie formy wyrazowej ściśle określonej przez użytkownika, przy jednoczesnym zachowaniu dowolności w zakresie form fleksyjnych oraz liczby elementów tworzących frazę. Warto zauważyć, że w omawianym oknie wyszukiwania można wprowadzić jednostkę językową zarówno w języku L1, jak i L2, co umożliwia wpisanie, np. w korpusie polskim, jednostki w języku polskim lub rosyjskim w zależności od potrzeb analizy, a następnie wygenerowanie listy konkordancji zawierających wskazaną jednostkę w jej kontekście tekstowym wraz z przekładem na drugi język. Poniżej przedstawiono formularz narzędzia *Поиск точных форм* wraz z przykładem konkordancji dla wyrażenia *jak śmiesz*:

Поиск точных форм ? ^

Слово или фраза

jak śmiesz

Искать

1. Л. Н. Толстой. Война и мир, том 2 (1865-1869) | Lew Tołstoj. Wojna i pokój, t. 2 (Andrzej Stawar, 1958) ⓘ

русский:

Как ты смеешь говорить, что он неблагороден? ⓘ ↔

польский:

Jak śmiesz mówić, że on jest nieszlachetny? ⓘ ↔

Kolejna funkcjonalność, znana również z innych korpusów, *Лексико-грамматический поиск*, pozwala na określenie bardziej szczegółowych kryteriów i parametrów dla zadanego zapytania, wśród których znajdują się:

- 1) *Лемма* – forma hasłowa wskazanego zapytania;
- 2) *Словоформа* – określona jednowyrazowa forma zapytania, jaka ma się pojawić w wynikach;
- 3) *Грамматические признаки* – możliwość określenia cech morfologicznych leksemu i/lub formy wyrazu;
- 4) *Семантика* – wybór konkretnych cech semantycznych i słowotwórczych leksemu i/lub formy wyrazu;
- 5) *Дополнительные признаки* – wyszukiwanie słów w określonej pozycji, zadanej przez wybrane parametry.

Omawiana możliwość, w odróżnieniu od poprzedniej, wymaga wyboru języka zbioru tekstowego, co oznacza, że np. po wybraniu opcji *только в польском* konieczne jest sformułowanie zapytania w języku polskim, a następnie funkcja wyszukuje odpowiadające mu tekstowe odpowiedniki w języku rosyjskim (i odwrotnie). Użytkownik może dodać dodatkowe słowo do procesu wyszukiwania, przy czym korpus wyszuka jedynie te teksty, w których oba słowa występują jednocześnie; w przypadku braku takiej sytuacji wyniki nie zostaną wyświetlone. Widok narzędzia, uzupełniony przykładem konkordancji dla jednostki *świat* w tłumaczeniu polsko-rosyjskim przedstawiono poniżej:

Лексико-грамматический поиск ? ^

Искать

☐ только в русском

☒ только в польском

+ Слово 1 +

Лемма

Словоформа

Грамм. признаки

Семантика

Доп. признаки

1. Wacław Radziwiłowicz. Nie mówcie nam: On się rozbił (Gazeta Wyborcza) (04.01.2011) | Вацлав Радзивиллович. Не говорите нам: «Он разбился» (Газета Выборча, 2011.01.04) (inosmi.ru, 2010) ⓘ

польский:

Wtedy — u schyłku lata 2000 r. — *świat* już wie, że na podkładzie "Kurska" była potworna eksplozja, ale Rosjanie mówią, że ich jednostkę staranował pewnie jakiś okręt zachodni. ☹️ ↔️

польский:

Trzeba publicznie, i to jak najszybciej, powiedzieć opinii publicznej w kraju i na *świecie*, jakie konkretnie zastrzeżenia mamy do roboty wykonanej przez MAK, na jakie konkretnie pytania nie odpowiadają nam Rosjanie. ☹️ ↔️

русский:

Тогда, на исходе лета 2000 года, мир уже знал, что на борту "Курска" произошел ужасный взрыв, но россияне говорили, что его наверняка протаранило какое-то западное судно. ☹️ ↔️

русский:

Следует публично и как можно скорее сообщить польской и мировой общественности, какие конкретно претензии есть у нас к работе, которую проделал MAK, на какие конкретно вопросы нам не отвечают россияне. ☹️ ↔️

Ostatnia funkcja, *Поиск на двух языках*, wykorzystuje interfejs identyczny jak w przypadku poprzednich zapytań, umożliwiając jednocześnie ustawienie dwóch kryteriów wyszukiwania dla każdego języka. Dzięki temu użytkownik otrzymuje fragmenty tekstu, w których zgodność z zapytaniami występuje zarówno w oryginale, jak i w tłumaczeniu. Przykład zastosowania danego parametru dla jednostki polskiej *kot* i rosyjskiej *собака* wraz z uzyskaną próbką wyników przedstawia się następująco:

Польский язык

Слово 1

Лемма

Словоформа

Грамм. признаки

Доп. признаки

Русский язык

Слово 1

Лемма

Словоформа

Грамм. признаки

Семантика

Доп. признаки

1. Andrzej Sapkowski. Boży bojownicy (1) (2004) | Анджей Сапковский. Божьи воины (1) (Е. Вайсброт, 2006) ⓘ

польский:

Dwa psy, jeden **kot** i ze trzydzieści gołębi w zgodzie i harmonii piło wodę z kałuży pod pręgierzem, nawet nie oglądając się na nielicznych przemykających pod ścianami domów przechodniów. ⓘ ↔

польский:

Przy kałuży u pręgierza pies się drapał, **kot** wylizywał sobie ogon i jego szeroko rozumiane okolice, gołębie dreptały skrajem wody. ⓘ ↔

польский:

Spod pręgierza uciekły nagle psy, za nimi pognął **kot**. ⓘ ↔

русский:

Две **собаки**, одна кошка и штук тридцать голубей в согласии и гармонии пили воду из лужи около позорного столба, не обращая внимания на немногочисленных прохожих, крадущихся вдоль стен домов. ⓘ ↔

русский:

У лужи при позорном столбе чесалась **собака**, кошка вылизывала хвост. Голуби топтались по краю воды. ⓘ ↔

русский:

От позорного столба вдруг сбежали **собаки**, за ними помчалась кошка. ⓘ ↔

Z kolei wyszukiwanie w korpusie wielojęzycznym odbywa się za pomocą interfejsu o zbliżonej strukturze, wzbogaconego o dodatkowe funkcje, umożliwiające wpisanie zapytania w dowolnym języku. Następnie korpus generuje odpowiednie konkordancje, przy czym język wyjściowy porównania zależy od języka oryginału, w którym znajduje się tłumaczenie wskazanego przez użytkownika leksemu. W polu *поиск точных форм* zapytanie analogicznie wpisywane jest w dokładnej formie, w jakiej ma wystąpić w tekście, przy czym język zapytania może być dowolny. Natomiast w polu *лексико-грамматический поиск* dodatkową funkcjonalnością jest parametr *перевод*, który umożliwia wyszukiwanie słownikowe leksemu oraz jego homonimów w określonym kierunku tłumaczenia. Po zadaniu zapytania użytkownik otrzymuje tekst wyjściowy w języku angielskim oraz tekst porównawczy w języku, w którym sformułowano zapytanie, z możliwością natychmiastowego wyboru tłumaczenia na inny język. Poniżej przedstawiono fragment pojedynczej konkordancji dla wcześniejszego zapytania *kot* oraz dodatkowe okno umożliwiające wybór innych przekładów, wraz z informacją o tłumaczu tekstu:

английский:

Down, down, down. There was nothing else to do, so Alice soon began talking again. "Dinah'll miss me very much to-night, I should think! "(Dinah was the cat.) "I hope they'll remember her saucer of milk at tea-time. Dinah, my dear! I

польский 1 (Переводчик: Antoni Marianowicz) ▾

W dół, w dół, wciąż w dół. Nie było nic do roboty, więc Alicja zabawiła się nadal rozmową z samą sobą, "Jacek będzie tęsknił za mną dziś wieczorem". (Jacek był to **kot**). "Mam nadzieję, że w domu nie zapomną dać mu mleka na



4.4.3 Anotacja w Korpusie Równoległym NKJR

W ramach dwujęzycznych par większość tekstów równoległych w NKJR jest wyrównywana za pomocą algorytmu programu *HunAlign*, który umożliwia ręczną korektę przy użyciu interfejsu *Euclid*. Dana procedura rozpoczyna się od wczytania plików tekstowych w formacie ANSI lub Unicode, które przechodzą automatyczną wstępną obróbkę, obejmującą m.in.: usunięcie pustych linii, nadmiarowych spacji oraz przenoszenie każdego zdania do nowej linii. Następnie program *HunAlign* wykorzystuje mechanizmy statystyczne, takie jak długość zdań, dopasowanie symboli czy struktura znaków, aby automatycznie przypisać każdej parze zdań współczynnik prawdopodobieństwa wyrównania, a fragmenty o niskim współczynniku są oznaczane jako wymagające ręcznej weryfikacji (Сичинава 2015: 211). Z kolei do przetwarzania dwóch równoległych tekstów – oryginału i tłumaczenia – dla korpusu angielskiego i niemieckiego wykorzystuje się inny program – *ПарТекст*. Wyrównywanie odbywa się w tym przypadku na poziomie zdań, co skutkuje stworzeniem tekstu syntetycznego, w którym każdemu zdaniu oryginału odpowiada przyporządkowane zdanie tłumaczenia. Algorytm segmentacji tekstu na zdania w obu narzędziach opiera się na prostych regułach wykorzystujących znaki interpunkcyjne, pomijając bardziej złożone struktury językowe, takie jak inicjały, skróty czy zasady mowy bezpośredniej (Сичинава 2012: 272). Jednakże zarówno w jednym, jak i w drugim programie często występuje problem tzw. wolnego tłumaczenia, który podczas analizy powoduje trudności związane m.in. z podziałem przez tłumacza jednego

zdania oryginalnego na dwa, zmianą części mowy wyrażającej podmiot oraz innymi modyfikacjami strukturalnymi tekstu (Добровольский, Кретов, Шаров 2005: 265).

Metainformacje są przechowywane w oddzielnym pliku w formacie CSV, w którym wartości komórek są oddzielone za pomocą znaku „przecinek-kropka”. Zestaw anotacji zewnętrznej obejmuje m.in. takie informacje, jak: tytuł tekstu w oryginale, rok jego powstania, imię autora i rok jego urodzenia, tytuł tekstu w tłumaczeniu, imię tłumacza, data przekładu, a także informacje o cechach gatunkowych tekstu, takich jak literacki czy publicystyczny. Dodatkowo w tabeli odnotowywany jest język oryginału oraz język tłumaczenia. Użytkownik może uzyskać dostęp do tych informacji poprzez kliknięcie w nagłówek otrzymanej konkordancji, zawierającej tekst równoległy odpowiadający zadanemu zapytaniu. Np. dla wcześniej wspomnianego zapytania *jak śmiesz* dane anotacji zewnętrznej z wyodrębnionej konkordancji przedstawiają się następująco:

Л. Н. Толстой. Война и мир, том 2 (1865-1869) Lew Tołstoj. Wojna i pokój, t. 2 (Andrzej Stawar, 1958)		×
Автор	Л. Н. Толстой	
Год рождения	1828	
Название	Война и мир, том 2	
Перевод имени автора	Lew Tołstoj	
Перевод названия	Wojna i pokój, t. 2	
Дата создания	1865-1869	
Переводчик	Andrzej Stawar	
Год перевода	1958	
Сфера функционирования	художественная	
Предложений	15100	
Словоформ	231700	
Язык оригинала	русский	
Язык перевода	польский	
		Сообщить об ошибке

Z kolei schemat anotacji wewnętrznej (morfologicznej) opiera się na jednolitym zestawie kategorii części mowy i odpowiadających im tagów gramatycznych, a współczesne zasady tworzenia korpusu zakładają konsekwentne i spójne oznaczanie tych kategorii w różnych językach, co umożliwia ujednolicenie analizy międzyjęzykowej. Wspomniany program Euklid umożliwia dodanie metaparametrów, takich jak dane o autorze, dacie czy językach, które są zapisywane w edytowalnym pliku tabelarycznym, co upraszcza zarządzanie i przetwarzanie danych korpusu. Teksty dla większości języków zostały oznakowane z wykorzystaniem specjalnych wersji programu *MyStem*, dostosowanego do analizy morfologicznej każdej z wymienionych baz językowych. Program ten umożliwia precyzyjne

oznaczanie form gramatycznych oraz usuwanie problemów związanych z homonią, np. w przypadku języka polskiego zastosowano algorytm *TaKiPi*, który został opracowany specjalnie z myślą o unikalnych cechach morfologii polszczyzny. Z kolei teksty w języku rosyjskim są przetwarzane z wykorzystaniem automatycznej morfologicznej analizy, która uwzględnia rozwiązywanie problemów związanych z homonią. Ponadto, teksty te poddawane są analizie semantycznej, której metodologia odpowiada semantycznemu oznakowaniu leksyki rosyjskiej w innych korpusach językowych. Wszystkie teksty w korpusie można zapisywać w formacie XML, co umożliwia ich ostateczne przechowywanie lub dalszą edycję, przy czym kluczowe informacje, takie jak język oryginału i tłumaczenia, są zapisywane w formie znaczników XML, a pary zdań otrzymują unikalne numery. Schemat anotacji wewnętrznej dla każdego słowa wchodzącego w skład uzyskanej pary konkordancji równoległej jest zróżnicowany i zależy od stopnia szczegółowości anotacji danego języka, przy czym wiadomo, że język rosyjski jako język wyjściowy dostarcza najpełniejszych metainformacji. Na podstawie dokonanych obserwacji można stwierdzić, że do charakterystyk stałych w opisie słowa należą:

- 1) *Лемма* – forma słownikowa wyrazu z możliwością odesłania do słownika w celu uzyskania pełnej informacji semantycznej konkretnej jednostki;
- 2) *Грамматика* – charakterystyka gramatyczna wyrazu;
- 3) *Дополнительные признаки* – informacja o pozycji wyrazu w zdaniu, opisana za pomocą tagów, dla których wyjaśnienie można znaleźć na stronie <https://ruscorpora.ru/page/annotation> [data dostępu: 04.01.2025].

Poniżej przedstawiono przykładowe wyniki uzyskane na podstawie losowo wybranych słów (*nieszlachetny* / *неблагороден*), mające na celu ukazanie wewnętrznej metainformacji. Można zauważyć, że w przypadku analogicznego słowa w języku rosyjskim schemat anotacji wewnętrznej został uzupełniony o dodatkową informację semantyczną:

nieszlachetny		×
Лемма	nieszlachetny (см. портрет слова, см. словарь)	
Грамматика	прилагательное, единственное, именительный, мужской, личное (pl)	
Доп. признаки	bques, bmark, last, numred, casered	

неблагороден		×
Лемма	неблагородный (см. портрет слова, см. словарь)	
Грамматика	прилагательное, краткая форма (ru), мужской, единственное	
Семантика	отадъективные, свойство человека, качества человека, качественные	
Доп. признаки	bques, bmark, last, genderred, numred, norm	

4.4.4 Zalety i ograniczenia Korpusu Równoległego NKJR

Rozwój równoległych korpusów językowych po stabilizacji nowej wersji wyszukiwarki Yandex dla NKJR zakłada przede wszystkim udoskonalanie istniejących par językowych, np. poprzez wzbogacenie korpusu o bułgarsko-rosyjskie teksty, które obecnie są reprezentowane wyłącznie w kierunku rosyjsko-bułgarskim, a także uzupełnienie zbiorów o teksty pozaliterackie w językach, gdzie są one obecnie nieobecne, takich jak ormiański czy litewski. Ponadto planowane jest pełne wdrożenie funkcji wyszukiwania według wszystkich dostępnych parametrów, w tym opcji wyszukiwania na podstawie słów kluczowych oraz tłumaczeń niedokładnych, jak również integracja narzędzi wizualizacyjnych, umożliwiających generowanie wykresów podobnych do tych wykorzystywanych w korpusach prasowych NKJR (Сичинава 2019: 57).

Do głównych zalet korpusu należy jego imponująca wielkość w odniesieniu do wybranych języków, takich jak angielski i niemiecki, co pozwala na realizację kompleksowych badań w obszarze konkretnych języków. Ważnym atutem jest także intuicyjny interfejs użytkownika, który znacząco ułatwia pracę z korpusem, umożliwiając jednocześnie formułowanie wieloskładnikowych zapytań. Dzięki możliwości dostosowania kryteriów wyszukiwania do różnych potrzeb badawczych, użytkownik ma sposobność precyzyjnego określenia parametrów, takich jak zakres, charakter czy typ zapytania. Rozwiązanie takie sprzyja uzyskiwaniu szczegółowych i wiarygodnych wyników, które są niezbędne przy analizie interesujących użytkownika ekwiwalentów tłumaczeniowych. W efekcie korpus staje się nie tylko narzędziem do prostych wyszukiwań, lecz także zaawansowaną platformą wspierającą badania nad kolokacjami, strukturami gramatycznymi czy różnicami semantycznymi między językami, co czyni go niezwykle uniwersalnym narzędziem w analizach kontrastywno-porównawczych. Jednym z innowacyjnych rozwiązań korpusu równoległego NKJR jest możliwość uchwycenia różnic między tekstem oryginalnym a jego tłumaczeniem dzięki zastosowaniu specjalnych atrybutów, takich jak *loose* czy *omit* (Сичинава 2013: 78), które dokumentują istotne zmiany – funkcjonalność ta często nie występuje w innych korpusach. Takie rozwiązanie jest szczególnie przydatne w analizie przekładu artystycznego, gdzie często obserwuje się znaczną swobodę w podejściu tłumacza. Przejawia się ona w różnorodnych modyfikacjach tekstu, takich jak pominięcia, dodania, zmiana kolejności zdań, a nawet całkowita zmiana znaczenia. Zabiegi te mogą być wynikiem twórczego przekształcenia tekstu przez tłumacza, który stara się oddać nie tylko treść, ale także styl i emocje oryginału, lub efektem ingerencji zewnętrznych, takich jak cenzura.

Jednym z podstawowych ograniczeń korpusu jest jego oparcie o teksty literatury pięknej, co w ograniczonym stopniu umożliwia analizę języka w rejestrach mniej formalnych, takich jak język potoczny (Трофимов 2018: 386). Wąska reprezentacja tekstów pochodzących z codziennej komunikacji, mediów społecznościowych czy dokumentów użytkowych znacząco ogranicza możliwości badania współczesnych zjawisk językowych, takich jak kolokwializmy, slang czy specjalistyczna terminologia używana w różnych środowiskach zawodowych. W konsekwencji korpus, choć bardzo wartościowy w kontekście analizy literackiej i formalnej, nie w pełni odzwierciedla pełne spektrum użycia języka w różnych sytuacjach komunikacyjnych. Szczególnym wyzwaniem w analizie kontrastywnej języków rosyjskiego, polskiego i czeskiego pozostaje stosunkowo ograniczona liczba dostępnych zasobów tekstowych w istniejących korpusach. Taka sytuacja znacząco utrudnia precyzyjne wyodrębnianie potencjalnych ekwiwalentów przekładowych dla analizowanych jednostek wielowyrazowych emotywnych i fatycznych. Rozbudowa tych korpusów o bardziej zróżnicowane i obszerne dane mogłaby znacząco poprawić efektywność analiz oraz umożliwić identyfikację bardziej precyzyjnych zależności między strukturami językowymi. Warto również zauważyć, że omawiany korpus (podobnie jak InterCorp) nie dysponuje funkcją biKWiC, która pozwalałaby na identyfikację równoważników słów kluczowych w różnych językach. Ogranicza to znacząco możliwość prowadzenia precyzyjnych wielojęzycznych analiz porównawczych, a implementacja takiego narzędzia mogłaby istotnie podnieść wartość analityczną korpusów. Repertuar anotacji wewnętrznej dla pozostałych języków również wymaga uzupełnienia, co było zauważalne w dysproporcji pomiędzy bardziej rozbudowaną metainformacją dostępną dla języka rosyjskiego a mniej szczegółową dla języka polskiego. Ten brak równowagi wskazuje na konieczność wprowadzenia poprawek, gdyż jedynie spójne i kompleksowe rozszerzenie anotacji dla wszystkich języków pozwoli na zapewnienie wysokiej precyzji analiz oraz zwiększenie ich wiarygodności i użyteczności.

ROZDZIAŁ V

Geneza pojęcia reproduktu. Inspiracje i rozwój myśli opolskiej

Niniejszy rozdział stanowi rozwinięcie myśli opolskiej dotyczącej pojęcia reproduktu, zaprezentowanego wstępnie w rozdziale drugim, a także omówienie obiektów badań, tj. fatyzmów i emotywów. W poprzednim rozdziale przedstawiono ogólne porównanie terminu *frazem* w rozumieniu Wojciecha Chlebdy z innymi zbliżonymi pojęciami językoznawczymi, w tym z teoriami badaczy czeskich i rosyjskich. Natomiast tu idea *reproduktu* zostanie poddana szczegółowej analizie, ukazującej jej specyfikę, konceptualne założenia oraz jej powiązania z innymi teoriami językoznawczymi.

Wspomniany wcześniej Andrzej Bogusławski w swoich badaniach nad językiem przyjął nowatorską perspektywę, kwestionując tradycyjne postrzeganie języka jako systemu opartego wyłącznie na pojedynczych wyrazach. W jego ujęciu podstawową kategorię analizy lingwistycznej stanowią jednostki języka, które mogą mieć zarówno postać jednowyrazową, jak i wielowyrazową, a różnica między nimi ma charakter czysto graficzny i nie wpływa na ich funkcjonalność w systemie komunikacji.

Tego rodzaju sposób myślenia o języku miał jednak swoich prekursorów w polskiej lingwistyce XX wieku. Stanisław Skorupka już w latach 40. XX wieku rozwijał koncepcję frazeologii obejmującą nie tylko tradycyjnie rozumiane związki stałe i łączliwe, ale także tzw. związki luźne – doraźnie tworzone połączenia wyrazowe, które choć pozornie swobodne, w rzeczywistości podlegają określonym regułom językowym (Skorupka 1969: 221–223). Tym samym podważył intuicyjne przekonanie, że w języku istnieją kombinacje całkowicie dowolne, torując drogę do późniejszych rozważań na temat wielowyrazowych jednostek językowych. Patrząc na jego dorobek z perspektywy kilkudziesięciu lat, można uznać go za prekursora idei formułiczności języka, czyli przekonania, że nasza komunikacja jest w znacznej mierze oparta na gotowych wzorcach czy strukturach frazeologicznych.

Z kolei Witold Doroszewski już w 1954 roku, publikując *Zagadnienia leksykografii polskiej*, pisał, że „słownik jest zbiorem wyrazów i takich połączeń wyrazowych, które są jednostkami leksykalnymi, to znaczy pełnią takie same funkcje jak wyrazy pojedyncze, a więc są jako całości znakami odrębnych desygnatów. Pod tym względem między wyrazami a związkami wyrazowymi zasadniczej różnicy nie ma.” (Doroszewski 1954: 85). Podobieństwo między jednostkami języka u Bogusławskiego a jednostkami leksykalnymi u Doroszewskiego

jest wyraźne – w obu przypadkach chodzi o takie twory werbalne, które niezależnie od swojej formy stanowią całościowe znaki odrębnych desygnatów, czyli innymi słowy, werbalizatory kwantów sensu i potencjałów treściowych.

Jednym z kluczowych zagadnień, jakie podjął Bogusławski, jest rola frazeologii w językoznawstwie, którą określał jako *centralną kwestię lingwistyczną*. W swojej pracy wskazywał, że zadaniem językoznawstwa jest nie tylko analiza pojedynczych wyrazów, lecz przede wszystkim identyfikacja jednostek językowych w ich rzeczywistej postaci, co oznacza konieczność tropienia i wyodrębniania jednostek wielowyrazowych. Jak pisał:

„Decydując się na powiązanie terminu „frazologia” z pojęciem „jednostki języka” [...] czynimy z problemów frazeologii po prostu centralną kwestię lingwistyczną. Jakkolwiek bowiem dokładnie scharakteryzujemy „jednostkę języka” [...] pojęcie to jest absolutnie konieczne i absolutnie fundamentalne; ponieważ zaś nie ma żadnych zewnętrznych wskaźników granic jednostek języka, a mogą one być w miejscach najprzeróżniejszych, przy czym mają tendencję do tego, by się właśnie nie pokrywać z granicami pojedynczych wyrazów graficznych, fonologicznych czy fleksyjnych, prawdziwym, pierwszoplanowym problemem badawczym tam, gdzie chodzi o to właściwe, poprzedzające wszystkie inne, zadanie językoznawstwa (w węższym sensie), jakim jest rejestracja jednostek języka, staje się analiza połączeń wyrazowych zmierzająca do oddzielenia regularnych produktów od quasi-kombinacji, tzn. faktycznych jednostek języka[...]” (Bogusławski 1989: 18–19).

Brak jednoznacznych wskaźników delimitujących początek i koniec jednostek wielowyrazowych sprawia, że ich wyszukiwanie i wyodrębnianie nie jest procesem automatycznym i, jak zauważa Bogusławski, wymaga nie tylko intuicyjnej znajomości języka, ale także zastosowania zaawansowanych metod analitycznych. Podkreśla on również, że skala występowania jednostek wielowyrazowych jest nieporównywalnie większa od liczby jednostek jednowyrazowych – ich liczba sięga grubych milionów, co czyni je dominującym elementem struktury języka. Założenia sformułowane przez Bogusławskiego stały się inspiracją dla wielu badaczy zajmujących się frazeologią, w tym dla Wojciecha Chlebdy i badaczy z ośrodka opolskiego. Koncepcja rozwijana w tradycji opolskiej opiera się na założeniu, że zasoby języka składają się zarówno z jedno-, jak i wielowyrazowych jednostek, które przeciwstawia się dynamicznie tworzonemu produktowi językowemu powstającemu w wyniku kontekstowego łączenia tychże elementów w procesie komunikacji. To podejście, zapoczątkowane w latach 70. przez Bogusławskiego, stanowi fundament dla późniejszych badań nad metodami identyfikacji, semantyzacji i kodyfikacji jednostek leksykalnych – zarówno z użyciem technik manualnych, jak i automatycznych. W konsekwencji język należy postrzegać nie tylko jako zamknięty system stałych elementów, lecz także jako przestrzeń generowania nowych struktur, zależnych od warunków komunikacyjnych. Na tej samej tradycji teoretycznej opierają się

również późniejsze propozycje terminologiczne Chlebdy dotyczące wielowyrazowych struktur językowych, w tym koncepcja *frazemu* i *reproduktu*, które rozwijają ideę języka jako systemu zawierającego zarówno stabilne jednostki, jak i elastyczne, kontekstowo konfigurowane struktury.

Wprowadzenie kategorii *frazemów* i *reproduktów* miało na celu ustanowienie nadrzędnego hiperonimu obejmującego zarówno klasyczne frazeologizmy, jak i ustabilizowane ciągi wielowyrazowe, niezależnie od ich dotychczasowej przynależności do tradycyjnie rozumianej frazeologii. W skład tej kategorii wchodzi nie tylko kanoniczne wyrażenia, takie jak *mieć z kimś na pieńku*, ale także skrzydlate słowa (*Jestem za, a nawet przeciw*), nazwy własne (*Wyspy Japońskie*), napisy miejskie (*świeżo malowane*), emotywy (*a niech to lich!*), fatyzmy (*moim zdaniem...*), zwroty etykiety językowej (*bardzo mi miło*), a także wszystkie pozostałe powtarzalne i utrwalone formacje językowe. Jednostki te, choć powszechnie funkcjonujące w komunikacji, były dotąd marginalizowane w leksykografii lub uznawane za mniej istotne ze względu na ich nietypową strukturę i brak zgodności z tradycyjnymi kryteriami klasyfikacji frazeologicznej. Założenia te wpisują się również w fundamentalną myśl twórcy teorii *sens–tekst*, I. Mielczuka, zgodnie z którą „ludzie nie mówią słowami, ludzie mówią frazemami” (Mielczuk 1995: 167–232), czyli gotowymi, odtwarzalnymi strukturami językowymi. Jego podejście, kładące nacisk na reprodukowalność jednostek językowych, znajduje wyraźne paralele w koncepcji *reproduktu* Wojciecha Chlebdy. Zarówno Mielczuk, jak i przedstawiciele szkoły opolskiej podkreślali, że struktury językowe nie funkcjonują w oderwaniu od schematów, lecz są kształtowane poprzez powtarzalne wzorce, które organizują procesy komunikacyjne i warunkują użycie jednostek w języku. Opolska koncepcja frazemów i reproduktów wyraźnie rozwija tę ideę, wskazując, że to nie pojedyncze słowa, lecz utrwalone struktury frazematyczne odgrywają kluczową rolę w językowej konceptualizacji rzeczywistości oraz w mechanizmach przetwarzania i interpretacji wypowiedzi. Choć Mielczuk odnosił swoją tezę do poziomu *parole*, nie przeczy to założeniom Andrzeja Bogusławskiego, który wskazywał, że tego rodzaju jednostki są również nieodłącznym elementem systemu języka (*langue*). W konsekwencji, by stworzyć językowy system komunikacyjny, jednostki te muszą skądś pochodzić, a ich źródłem jest właśnie strukturalna organizacja języka, oparta na konstrukcjach frazematycznych. Podobne ujęcie znajduje się w koncepcji wspomnianego uprzednio Františka Čermáka, który określa frazem jako niemodelową jednostkę językową o ograniczonej możliwości modyfikacji składników. Ów badacz również wskazuje na reprodukowalność i stabilność jednostek frazematycznych, ale jego kluczowym kryterium pozostaje ich anomalia – brak możliwości swobodnej zamiany składników. Zasadnicza różnica

między ujęciem Čermáka a koncepcją Chlebdy sprowadza się do kryterium strukturalnego – w pierwszym przypadku frazem stanowi połączenie co najmniej dwóch elementów językowych, natomiast w podejściu drugiego za frazemy uznaje się również jednostki jednowyrazowe. Czeski badacz twierdził, że do frazeologii należą przede wszystkim nieprawidłowe kombinacje wyrazowe (*anomální kombinace*). Przykładem może być frazem *ležet polem*, który stanowi konstrukcję formalnie nietypową: nie istnieją bowiem w języku inne analogiczne zestawienia statycznego czasownika *ležet* z rzeczownikiem *pole*. O anomalności decyduje tu również poziom semantyczny, tj. oba komponenty łączą się bowiem w sposób całkowicie odmienny niż w regularnej kolokacji *ležet na poli*, gdzie czasownik współwystępuje z określeniem adverbialnym. Čermák wskazuje także, że anomalne kombinacje mogły powstawać diachronicznie na różne sposoby, szczególnie poprzez porównania oraz metafory indywidualne. Wiele z nich stanowi dziś pozostałość dawnych możliwości kombinacyjnych, które współczesny system językowy już utracił. Ponadto, Čermák zwraca uwagę na funkcję nominatywną frazemów i idiomów. Ich szczególną właściwością jest to, że znoszą reguły obowiązujące w języku ogólnym: „co tu vypadá jako pád, funkčné pád není, co se zdá být objektem verba, jím není, slovo tu není slovem se svým povědomým významem, tvar slova není tvarem lexému stojícího v pozadí, protože ten tu není relevantní a jako celek sem vůbec nevstupuje apod.”. Podsumowując, teoria Čermáka, mimo że skupiona bardziej na idiomatyczności i stopniu frazeologizacji, wpisuje się w jednak ogólną ideę języka jako systemu opierającego się na utrwalonych wzorcach komunikacyjnych (Čermák 2007: 26–34).

Pomimo faktu, iż koncepcja frazemów rozwijała ideę języka jako systemu opartego na gotowych strukturach, nie wszyscy badacze uznali ją za wystarczająco precyzyjną, dlatego Wojciech Chlebda zdecydował się zrezygnować z tego terminu na rzecz bardziej jednoznacznego określenia – *reproduktu*. Dotychczasowe wykorzystywanie terminu *frazem* jako nadrzędnej kategorii obejmującej różnorodne pod względem formy, znaczenia i funkcji ciągi wyrazowe prowadziło do nieścisłości oraz problemów terminologicznych w ramach frazematyki, zwłaszcza że pojęcie to było już zakorzenione w innych koncepcjach językoznawczych i zaczęło funkcjonować w znaczeniach odbiegających od pierwotnych założeń. W rezultacie przyczyniło się to do terminologicznego chaosu, który skłonił badacza do poszukiwania bardziej precyzyjnego określenia. Nowy termin nie tylko akcentował reprodukowalność jako kluczową cechę analizowanych struktur, lecz także stanowił wyraźne przeciwstawienie wobec pojęcia *produktu językowego* w ujęciu Andrzeja Bogusławskiego, jednocześnie pełniąc funkcję środka werbalizacji określonego zasobu treściowego, obejmującego zarówno sądy i pojęcia, jak i intencje oraz emocje. Podkreślał również

przynależność do struktury językowej jako elementu systemu lub jego podsystemu, a zarazem był traktowany jako gotowy schemat językowy, który użytkownik aktywuje automatycznie w określonych kontekstach komunikacyjnych. Przykładem ich leksykograficznego opracowania stał się opolski *Idiomykon* – innowacyjny model słownikowy, który dokumentuje wielowyrazowe struktury, nadając im status pełnoprawnych jednostek języka poprzez określenie ich funkcji i przypisanie ekwiwalentów rosyjskich. Jednak wyznaczenie wyraźnej granicy między klasycznymi frazeologizmami a jednostkami niestandardowymi, takimi jak frazemy i reprodukt, nadal pozostaje problematyczne, gdyż kategorie te płynnie się przenikają. Kluczowym staje się więc precyzyjne zdefiniowanie pojęcia *reproduktu*, jego struktury i miejsca w systemie językowym.

5.1. Reprodukt

W celu ukierunkowania rozważań na samo pojęcie *reproduktu* oraz wyznaczenia punktu odniesienia dla dalszej analizy, zasadnym wydaje się ponowne przytoczenie definicji Wojciecha Chlebdy, która przedstawia się następująco: „Przez reprodukt należy rozumieć jednostkę języka (składnik systemu danego języka etnicznego), wyodrębnioną z tekstów, sformułowanych w tym języku, w rezultacie stwierdzenie jej regularnej powtarzalności w tych tekstach w funkcji werbalizatora tego, „co autor chciał powiedzieć”, czyli pewnej wiązki sensów, intencji, emocji, określonego zespołu tekstowego” (Chlebda 2010b: 15). Powszechne występowanie w mowie i piśmie wielowyrazowych oraz reprodukowalnych werbalizatorów pojęć, takich jak *na moje oko*; *co my tu mamy...*; *cała przyjemność po mojej stronie*; *wstęp wzbroniony*; *panie proszą panów* oraz wielu innych, nie ulega wątpliwości i świadczy o ich zakorzenieniu w systemie językowym jako ustabilizowanych jednostek komunikacyjnych, pełniących funkcję nośników określonych treści semantycznych i pragmatycznych. Reprodukty można zatem definiować jako zbiór ustabilizowanych wielowyrazowych całości nominatywnych, które pod względem formy i znaczenia funkcjonują jako spójne ciągi werbalne, tworząc fundamentalną warstwę języka i wypełniają przestrzeń systemu językowego. W określonych sytuacjach komunikacyjnych są one przywoływane z pamięci i odtwarzane w ustalonej postaci, co sugeruje, że w analogicznych warunkach użytkownicy innych języków również posługują się gotowymi strukturami leksykalnymi do wyrażania określonych pojęć, ocen czy emocji. Wojciech Chlebda określa ten mechanizm mianem *pamięci potocznej*, podkreślając jej kluczową rolę w procesie nauki języka – opanowanie tych jednostek w ich utrwalonej, konwencjonalnej formie jest niezbędne dla skutecznej komunikacji.

W rozumieniu przytoczonej definicji reproduktu określone teksty, jako specyficzne wypowiedzi jednostkowe, odznaczają się autorskim i niepowtarzalnym charakterem, jednak ich definicyjna powtarzalność wskazuje na stopniowe uniezależnianie się od ich pierwotnego twórcy. Należy również podkreślić, iż danego rodzaju jednostki nie powstają od podstaw w procesie twórczym autora określonej wypowiedzi, lecz zostają przez niego przejęte i włączone do jego komunikatu w konkretnej sytuacji komunikacyjnej. W rezultacie nie są autonomicznym wytworem pojedynczego twórcy, lecz przejawem mechanizmu odtwarzania i adaptacji wcześniej ukształtowanych struktur językowych, które miały swoje źródło w innych ramach czasowych czy komunikacyjnych. Innymi słowy, określone jednostki zaczynają funkcjonować jako byty niezależne w szerszym obiegu językowym, co oznacza, że stanowią element wspólny dla nieokreślonej liczby tekstów różnych autorów, a tym samym ich obecność i utrwalenie w komunikacji sugeruje ich zakorzenienie w powszechnej praktyce językowej. Wojciech Chlebda określił tę procedurę mianem przejmowania reproduktu, rozumianego jako działanie realizowane w przestrzeni napięcia pomiędzy dwoma biegunami:

1. **Autor A** zapożycza się u **autora B**, przejmując pewien wycinek jego wypowiedzi i oznakowuje je stosownymi znakami praw autorskich, np. cytatem bądź wskaźnikiem, typu *jak powiedział B*;
2. **Autor A** przejmuje pewien ukształtowany ciąg werbalny z systemu swojego języka, nie znając i nie dociekając jego autorstwa, tym samym korzysta więc z zasobów swego języka tworzonych przez reprodukt jedno- i wielowyrazowe (Chlebda 2010b: 15).

Między dwoma wskazanymi biegunami zachodzi nieustanny ruch dwukierunkowy, w ramach którego reprodukt należący do pierwszej grupy, werbalizujący społecznie istotne sensy, może stopniowo ulegać przejęciu przez autorów najnowszych tekstów. Proces ten prowadzi do jego stopniowego odrywania się zarówno od pierwotnego kontekstu tekstowego, jak i od autora źródłowego, co skutkuje dezautoryzacją, a w efekcie reprodukt przekształca się w element dobra powszechnego, dostępnego ogółowi użytkowników języka. Reprodukty pierwszej grupy to w przeważającej mierze skrzydlate wyrażenia, które są już dobrze udokumentowane w językoznawstwie czy leksykografii oraz posiadają wyraźnie określonego autora. Z kolei myśl opolska jest ukierunkowana na reprodukt należące do drugiej grupy, które albo już uzyskały status jednostek systemu językowego, albo znajdują się w procesie jego nabywania, a które dotychczas były jedynie w niewielkim stopniu rejestrowane w opracowaniach leksykografii jedno- i wielojęzycznej.

Pod względem struktury reprodukt mogą przyjmować postać jedno- lub wielowyrazową. Reprodukty jednowyrazowe zostały już szczegółowo zbadane, a ich

wyodrębnienie w tekście jest naturalne ze względu na wyraźnie zaznaczone granice początku i końca wyrazu. W przeciwieństwie do nich reproduktów wielowyrazowe cechują się większą złożonością strukturalną i obejmują zarówno jednostki niezdaniowe, jak i zdaniowe, a także zwarte konfiguracje zdań. Właśnie na tej kategorii – reprodutach wielowyrazowych – koncentrują swoje badania Wojciech Chlebda oraz zespół opolski, dostrzegając w nich dotychczas niedostatecznie opisany obszar analiz lingwistycznych:

1. Reprodukty leksykalnie zamknięte cechują się ustalonym układem składników leksykalnych, który pozostaje w dużej mierze niezmienny i podlegają w wypowiedzi odtworzeniu, np. *czarna dziura; gruba kreska; żelazna kurtyna; zimna wojna; efekt motyla; droga mleczna* itp.

2. Reprodukty leksykalnie półotwarte – charakteryzują się strukturą, w której część komponentów leksykalnych jest niezmienna i narzucona przez system językowy, natomiast pozostałe elementy mają określone miejsce lub miejsca w strukturze reprodutu, lecz ich konkretna realizacja leksykalna zależy od kontekstu realizacji wypowiedzi. Współczesna leksykografia przyjęła konwencję zapisu, według której stałe składniki reprodutów wyróżnia się półtłustym drukiem, natomiast elementy zmienne – jasną kursywą, np. *Kto **siedzi u kogo w kieszeni**; Kto ma haka na kogo; Co stoi na przeszkodzie czemu* itp.

3. Reprodukty leksykalnie otwarte – których wszystkie składniki leksykalne pozostają nieokreślone w formie zapisu, czyli występują *in absentia*. Oznacza to, że w reprezentacji pisemnej nie pojawiają się żadne konkretne elementy leksykalne, a jedynie abstrakcyjny schemat konstrukcyjny. W związku z tym, w procesie realizacji językowej użytkownik musi samodzielnie uzupełnić strukturę odpowiednimi wyrazami, zgodnie z wymaganiami kontekstu wypowiedzi. Tym samym reprodukowana jest nie konkretna treść, lecz ogólny wzorzec umożliwiający tworzenie licznych wariantów danej konstrukcji, np. *Coś-, czyjeś-; Ktoś-/coś-, czegoś-* itp.

W opracowaniach leksykografii jedno- czy dwujęzycznej dominują głównie reproduktów leksykalnie zamknięte oraz te z ogólnie określoną walencją, podczas gdy reproduktów półotwartych i otwartych występują znacznie rzadziej lub w ogóle nie są uwzględniane. Repetytywność danych jednostek sprawia, że określone wielowyrazowce, poprzez częste występowanie w określonych kontekstach, stopniowo stają się integralnymi składnikami systemu językowego. Należy podkreślić, iż analiza danych jednostek nigdy nie odbywa się w oderwaniu od konkretnej

sytuacji komunikacyjnej, lecz zawsze winna uwzględniać występujący w komunikacie kontekst. Takie podejście jest w pełni uzasadnione, gdyż myśl Wojciecha Chlebda, czerpiąca z koncepcji frazeologii nadawcy, czyli frazematyki, zakłada, że wszystkie jednostki językowe powinny być rozpatrywane przede wszystkim z perspektywy swojej intencjonalności. W tym kontekście kluczową rolę odgrywa ramka pragmatyczna (Chlebda 2020: 35), która stanowi podstawę dla analiz porównawczych prowadzonych w różnych obszarach wypowiedzi, umożliwiając nie tylko dokładniejszą charakterystykę badanych struktur, lecz także uchwycenie ich funkcji w rzeczywistych aktach komunikacyjnych:

[KTO] – [KIEDY] – [GDZIE] – [O CZYM] – [W JAKIM CELU] – [W JAKI SPOSÓB]

ZWRACA SIĘ

[DO KOGO]

[Z JAKICH PRZYCZYŃ]

[Z JAKIM SKUTKIEM]

Powyższy dziewięcioelementowy schemat stanowi zatem podstawowe narzędzie analizy frazematycznej, które może być stosowane poprzez „przykładanie” go do konkretnych aktów reprodukcji jednostek językowych, co pozwala na systematyczne badanie ich funkcjonowania w różnych kontekstach wypowiedzi. Dzięki temu możliwym staje się nie tylko uzyskanie pełnego obrazu danej jednostki poprzez odpowiedzi na pytania dotyczące nadawcy, miejsca i czasu użycia, formy jednostki, jej adresata, przyczyn i celu wypowiedzi oraz wywołanego skutku, lecz także odniesienie jej do sytuacji w innym języku w celu ustalenia analogicznych ekwiwalentów przekładowych.

Reprodukowalność oraz odtwarzalność stanowią więc definicyjne cechy reproduktów, jednak nie należy ich utożsamiać z mechanicznym kopiowaniem raz ustalonych wzorców. Należy podkreślić, iż proces reprodukcji nie sprowadza się do biernego powielania identycznych kopii, gdyż każdy akt użycia języka dokonuje się w unikalnym kontekście sytuacyjnym i w związku z tym każda realizacja reproduktu zawiera element twórczy, polegający na jego dostosowaniu do nowej sytuacji komunikacyjnej. Proces ten może prowadzić do innowacyjnych modyfikacji, obejmujących zmiany strukturalne reproduktu, wynikające z jego dynamicznego dostosowywania się do aktualnego aktu komunikacyjnego. Innymi słowy, powielanie reproduktów charakteryzuje się często nieuniknioną zmiennością ich formy i znaczenia, wynikającą z dwóch głównych czynników. Po pierwsze, błędy w

reprodukcji są naturalnym efektem niedokładności powtarzania, która może wynikać z cech indywidualnych użytkowników języka oraz zmienności zachodzących kontekstów komunikacyjnych. Po drugie, reproduktory są przechowywane w pamięci użytkownika w dwóch postaciach: jako całościowe jednostki leksykalne o ustalonym znaczeniu oraz jako zbiór odrębnych komponentów, które mogą być interpretowane niezależnie od siebie. Ta dualistyczna natura rozpatrywanych jednostek sprawia, że mogą one być używane zarówno w formie niezmienniczej, jak i w sposób elastyczny, poprzez odwołanie do znaczeń ich poszczególnych elementów. Zjawisko to powoduje trudności kodyfikacyjne, związane z określeniem systemowej postaci reproduktu, czyli formy, która powinna zostać umieszczona jako artykuł hasłowy w słowniku. Wynika to z faktu, że jednostka występująca w rzeczywistych tekstach przybiera wiele różnych form kształtowych, co prowadzi do niejednoznaczności w jej klasyfikacji, a tym samym wymaga rozróżnienia między zakładaną postacią systemową reproduktu a jej wariantowością, co dodatkowo komplikuje proces leksykograficznego opisu, np. *Nie ma się czym przejmować* → *Nie ma sensu się tym przejmować* → *Nie warto się tym przejmować* itp.

Jednym z kluczowych, a zarazem najbardziej wymagających zagadnień było określenie podstawowych obszarów funkcjonowania reproduktów w powszechnym użyciu. Badacze z zespołu opolskiego, kierując się podziałem przyjętym w idiomatykonie, wyróżnili najczęściej występujące sfery ich zastosowania. Niemniej należy podkreślić, że klasyfikacja ta nie obejmuje wszystkich możliwych przypadków, gdyż reproduktory są obecne w niemal każdej dziedzinie ludzkiej aktywności. Z tego względu poniżej zaprezentowano jedynie najliczniejsze i najbardziej reprezentatywne obszary ich użycia, które jednocześnie wskazują charakterystyczne rodzaje tekstów, w jakich odtwarzalne jednostki wielowyrazowe występują szczególnie często. Przedstawiona poniżej klasyfikacja obejmuje wybrane kategorie z dziesięciu tomów idiomatykonu wraz z przykładami reproduktów zamkniętych i półotwartych:

Obszary użycia reproduktów	Przykładowe reproduktory
Dokumenty urzędowe i prawne (np. akty, umowy, zaświadczenia, deklaracje)	<i>zwolnienie lekarskie; Deklaracja Niepodległości; Sąd orzekł <u>co</u>; akt adopcji</i>
Teksty informacyjne w przestrzeni publicznej (np. napisy na tabliczkach, plakatach, apelach, ogłoszeniach)	<i>dozwolone od lat 18; parking płatny; nauka jazdy; uwaga zły pies</i>
Nazwy instytucji, urzędów i stanowisk itp.	<i>Głowa państwa; Agencja Bezpieczeństwa Wewnętrznego; samorząd terytorialny</i>
Historyzmy (np. nazwy wydarzeń, procesów społecznych, konfliktów)	<i>bitwa pod Warną; epoka żelaza; Komuna Paryska; monarchia absolutna; rewolucja lutowa</i>
Język codzienny i frazeologia (frazyzmy, emotywy, etykieta językowa itp.)	<i>bez pracy nie ma kołaczy; co nagle, to po diable; moim zdaniem....; co ty sobie myślisz?; przepraszam za spóźnienie</i>

Język cyfrowy i techniczny (np. komunikacja internetowa, interfejs użytkownika, terminologia IT)	<i>kod aktywacyjny; drukuj bieżącą stronę; konto dostępu; <u>kto</u> wysłał SMSa</i>
Nazwy własne (antroponimy, toponimy, tytuły itp.)	<i>Góry Sudety; Tańczący z wilkami; Adam Mickiewicz</i>
Terminologia naukowa i specjalistyczna (np. z dziedziny literatury, językoznawstwa, medycyny)	<i>spółgłoska bezdźwięczna; <u>co</u> się nie odmienia; akt illokucyjny; blok operacyjny; <u>co</u> zagraża zdrowiu</i>
Teksty reklamowe i marketingowe	<i>wielka wyprzedaż; ostatnie sztuki w magazynie; darmowa dostawa</i>
Teksty polityczne i publicystyczne	<i>dobra zmiana; głosowanie tajne; <u>kto</u> objął urząd</i>
Nazwy miar, jednostek ilościowych i zbiorczych	<i>kopa jaj; kanister benzyny; butelka mleka</i>
Nazwy produktów i artykułów konsumpcyjnych	<i>bułka tarta; kawa rozpuszczalna; sól do kąpieli, szampon przeciwłupieżowy</i>
Komparatywy (np. frazy porównawcze i wyrażenia idiomatyczne)	<i>jak <jasna> cholera; jak na zawołanie; jak malowany</i>
Język sportowy (np. nazwy dyscyplin i technik sportowych)	<i>hokej na trawie; wspinaczka górską;</i>
Terminologia kultury i sfery artystycznej	<i>gitara akustyczna; budowla eklektyczna; bilet wstępu</i>
Język biznesu i ekonomii	<i>obsługa klienta; krzywa dochodu; majątek bierny</i>
Nazwy aktywności człowieka	<i>napad smutku; praca na pełen etat; <u>kto</u> okazuje empatię <u>komu</u></i>

Jak pokazuje przedstawiona klasyfikacja, produkty funkcjonują w niemal każdej sferze użycia języka, od komunikacji urzędowej i specjalistycznej po codzienną mowę potoczną, co świadczy o ich kluczowej roli w organizacji wypowiedzi i kształtowaniu interakcji językowej. Nieprzypadkowo zatem w rozważaniach nad samym pojęciem produktu konieczne jest zwrócenie uwagi na ich podział i sfery występowania, gdyż prowadzi to bezpośrednio do dalszej analizy fatyzmów i emotywów – szczególnej kategorii jednostek, które, choć są nieodłącznym elementem żywej komunikacji, pozostają słabo przebadane. Język mówiony, w którym odgrywają one fundamentalną rolę, przez długi czas był marginalizowany w badaniach lingwistycznych i leksykograficznych, co skutkowało niedostatecznym uwzględnieniem wyrażen służących zarówno podtrzymywaniu kontaktu, jak i emocjonalnej reakcji na komunikaty rozmówcy. Fatyzmy, zapewniające płynność interakcji, oraz emotywy, będące spontanicznymi odpowiedziami wyrażającymi emocje, nie tylko organizują wypowiedź, ale także stanowią podstawę skutecznej komunikacji. Ich bogactwo i funkcjonalność znacząco wykraczają poza dotychczasowe opracowania. Z tego względu to właśnie fatyzmy i emotywy staną się przedmiotem szczegółowej analizy, której celem będzie ukazanie ich roli w strukturze języka oraz mechanizmów ich użycia w komunikacji.

5.2. Emotywy

Przez długi czas badacze pomijali emocje jako obiekt badań, jednak wraz z rozwojem podejścia antropocentrycznego zaczęto postrzegać człowieka jako centralny punkt odniesienia w języku, uwzględniając nie tylko opis cech otaczającej go rzeczywistości, lecz także procesy wewnętrzne, takie jak myśli, uczucia czy doznania. Językoznawcy zaczęli koncentrować się na językowych sposobach wyrażania stanów emocjonalnych, których ocena, kształtowana przez mechanizmy psychiczne, znajduje odzwierciedlenie w semantyce słów, wyrażeni i struktur składniowych, a następnie aktualizuje się w mowie. Emocje, jako wielofunkcyjny mechanizm psychiczny, pełnią kluczową rolę w procesach poznawczych, komunikacyjnych i interakcyjnych człowieka, wpływając zarówno na percepcję rzeczywistości, jak i sposób jej werbalizacji.

W tym kontekście warto zwrócić uwagę na ujęcie czeskich badaczy, zgodnie z którym „Slovo emoce se používá k označení pocitů a s nimi spojených myšlenek, psychických a fyzických stavů, a také řady pohnutek k určitému jednání. (Goleman 1997: 273) [...] V emoci zahrnuto kromě citu i chování, bezděčný mimický výraz a fyziologické dění” (Říčan 2009: 104).

I. Izard podkreśla, że emocje determinują ludzkie myśli (1999), co znajduje potwierdzenie w ujęciach Dodonowa (1978), Simonowa (1981) i Szingarowa (1971), którzy wskazują na ich wiele funkcji, m.in.: odzwierciedlającą, sygnalizacyjną, (auto) sugestywną, poznawczą czy oceniającą. Współczesna psychologia postrzega emocje jako złożone zjawisko psychiczne i behawioralne, mające kluczowe znaczenie dla motywacji i regulacji zachowania. Jak ujmują to czeska literatura psychologiczna: „Emoce jsou psychickým projevem, který má zásadní význam pro regulaci chování, neboť jsou podstatou jeho motivace (mají povahu pohnutek) i jeho organizace na základě individuálních zkušeností” (Bilík 2014: 7). Szczególnie istotna w kontekście języka i komunikacji jest funkcja poznawcza emocji, która nie tylko inicjuje procesy poznawcze, lecz także kompensuje braki w wiedzy i stymuluje ich rozwój (Шаховский 2009: 30). Współczesne badania nad relacją języka i emocji prowadzone są z różnych perspektyw, obejmujących aspekt strukturalno-semantyczny, stylistyczny, pragmatyczny, kognitywny, psycholingwistyczny czy lingwokulturowy. Rozwój badań doprowadził w konsekwencji do wyodrębnienia emotologii, czyli subdyscypliny zajmującej się werbalizacją, wyrażaniem i komunikacją emocji, która analizuje m.in. typologię znaków emocjonalnych, ich dynamikę, różnice w nominacji emocji w różnych językach oraz problemy związane z przekładem i lukami emocjonalnymi (Илинская 2006: 98). Zdecydowana

większość badań językoznawczych koncentruje się na analizie nacechowania emocjonalnego poszczególnych jednostek językowych, ich występowaniu w tekście oraz pełnionych przez nie funkcjach – na przykład w wypowiedziach typu *Co za piękny dzień* (o wartościowaniu pozytywnym) czy *Jestem dziś strasznie zmęczony* (o wartościowaniu negatywnym). Przedmiotem analizy są jednak przede wszystkim gotowe, reprodukowalne wypowiedzi emotywnie, które – choć rzadziej eksplorowane – nie zostały całkowicie pominięte w dotychczasowych badaniach językoznawczych.

Należy mieć na uwadze fakt, że ekspresja odnosi się do dynamicznej operacji językowej, czyli samego aktu wyrażania emocji w komunikacji. Ekspresywność zaś to relacyjna, konwencjonalnie utrwalona cecha jednostek językowych, będąca wynikiem tej operacji i możliwa do analizy naukowej (Skubalanka 1972: 124). Tak rozumiane ujęcie zbieżne jest z podejściem prezentowanym przez Marie Čechovą, która definiuje ekspresywność jako składnik znaczenia leksykalnego słowa, a jej charakter systemowy wyraźnie oddziela od aktualizacyjnych mechanizmów mowy:

„Expresivita slova je složka jeho významu, zahrnující citový a volní vztah ke skutečnosti, projevuje se na formě slova (expresivita inherentní), nebo až v kontextu (expresivita adherentní). Je-li expresivita složka jeho významu, patří expresiva do jazyka. Expresivita dosahovaná užitím (kontextem) patří do řeči: sloveso žrát ve spojení se zvířetem není expresivum, stává se jím, užije-li se ho o člověku” (Čechová i in. 2000: 63).

Ekspresywność języka jako jedna z podstawowych funkcji języka odnosi się zatem do zdolności nadawcy do werbalizacji swoich stanów intencjonalnych, zarówno emocjonalnych, jak i racjonalnych, co przejawia się w subiektywnym nacechowaniu wypowiedzi i ujawnia stosunek mówiącego do przekazywanej treści (Kiklewicz 2008: 17).

Zgodnie z koncepcją K. Bühlera znak językowy pełni w tym kontekście rolę symptomu wewnętrznego stanu nadawcy, a jego ekspresywna funkcja nie ogranicza się jedynie do przekazu informacji, lecz stanowi akt interpretacyjny, nadający komunikacji wymiar pragmatyczny i aksjologiczny (Bühler 2004: 29).

W literaturze przedmiotu ematywne akty mowy określane są nierzadko mianem wypowiedzi ekspresywnych, choć I. W. Trufanowa proponuje ich bardziej szczegółową klasyfikację, obejmującą ematywy (werbalizację osobistych emocji nadawcy), akty emotywno-oceniające oraz akty mowy ukierunkowane na emocjonalne oddziaływanie na odbiorcę (Труфанова 2000: 211–239). Poprawne użycie i interpretacja wypowiedzi ematywnych świadczą o kompetencji emocjonalnej użytkownika języka, a ich niewłaściwe stosowanie może prowadzić do zakłóceń w komunikacji, zarówno w obrębie jednej kultury, jak i w kontekście

międzykulturowym, co sprawia, że „ten sam” tekst w ustach różnych autorów może nabierać odmiennych znaczeń (Львовская 1985: 78). W procesie komunikacji uczestnicy nie tylko przekazują treści, lecz także dzielą się emocjami, wpływając na rozmówcę za pomocą środków językowych odzwierciedlających ich wewnętrzny stan i intencje. L. Kozuchowa podkreśla, że zwrot do rozmówcy, szczególnie na początku wypowiedzi, poprzedza właściwy komunikat i nadaje ton całemu procesowi porozumiewania się (Кожухова 2007: 83).

Emotywy, jako wyraz emocjonalnego nastawienia nadawcy wobec odbiorcy, odgrywają istotną rolę w budowaniu relacji międzyludzkich. Ich znaczenie często wykracza poza literalny sens wypowiedzi, co w kontekście komunikacji międzykulturowej może prowadzić do trudności interpretacyjnych oraz zakłóceń w porozumieniu (Солдатова 2009: 42). Wypowiedzi ematywne stanowią zatem jednostki językowe służące wyrażaniu emocji, uczuć i przeżyć mówiącego, najczęściej funkcjonujące w mowie potocznej, które charakteryzują się wyraźnym nacechowaniem subiektywnym oraz oparciem na utrwalonych w języku modelach składniowych. Struktura takich modeli zawiera zazwyczaj komponent leksykalny o stałym charakterze, zmienny element gramatyczny dostosowywany do kontekstu, a także fakultatywne składniki opcjonalne, które – choć nieobowiązkowe – wpływają na pełnię realizacji wypowiedzi w praktyce komunikacyjnej (Довгаль 2008: 112). Z uwagi na ekspresyjny charakter języka mówionego, materiałem badawczym w analizach często są dialogowe repliki postaci literackich bądź autentyczne nagrania wypowiedzi w języku mówionym (Czapiga 2015: 19).

Jak zauważa N. Ju. Borodulina, «Корпус оценочных единиц в языке велик и многообразен, так как оценка лиц, предметов, качеств, событий, осуществляемая субъектом, сопровождает процесс познания, проникает во все языковые ярусы и отражается средствами всех языковых уровней» (Бородулина 2007: 129–140). W każdej wypowiedzi współlistnieją propozycja, intencja oraz sens emocjonalny, tworząc triadę: oceny, emocjonalności i relacyjności.

Ocena – według G. W. Kołszanskiego – jest immanentną częścią aktu poznawczego, będącego efektem operacji myślowej nad przedmiotem wypowiedzi, i pojawia się zawsze tam, gdzie podmiot wchodzi w kontakt ze światem obiektywnym (Колшанский 1975: 142).

Jak zauważa G. N. Leńko, «оценка является необходимым компонентом эмоциональной реакции, так как определение значимости события или ситуации, а, следовательно, и активизации эмоции, происходит через оценивание» (Ленько 2015: 89).

Z kolei Sz. Bali już na początku XX wieku zauważał, że każdy akt mowy, nawet najbardziej neutralny zwrot, niesie ze sobą ładunek emocjonalny, a twierdzenia i zaprzeczenia

nigdy nie są całkowicie obiektywne – ich emocjonalne nacechowanie zależy od znaczenia, jakie przypisuje im mówiący (Балли 2003: 33). Tym samym, każda wypowiedź staje się potencjalnie nacechowana, a emocje nadawcy wpływają nie tylko na wybór środków językowych, ale i na formę całej wypowiedzi. Dodatkowo, istotną rolę odgrywają niewerbalizowane elementy oraz zdolność odbiorcy do ich uzupełniania, dzięki czemu wypowiedzi implicytne mogą niekiedy zawierać większy ładunek znaczeniowy niż wypowiedzi w pełni wyrażone (Трофимова 2008: 156).

Złożoność interpretacji emocjonalnej wypowiedzi można ująć w formie następującego cytatu z pracy Jana Bilíka: „Není tedy vždy správné určit právě jeden názor na analyzovaný text, ale jako vhodné se zdá určit názor či náladu z pohledu nejrůznějších čtenářů a zohlednit navíc i čas, což je podle mého názoru velmi složitý problém” (Bilík 2014: 3–4).

W idiomatykonach określa się wspomniane emotywy jako jednostki językowe pełniące funkcję spontanicznych replik w bezpośredniej komunikacji, stanowiące reakcję na wypowiedzi, gesty czy zachowania rozmówcy i służące wyrażeniu szerokiego spektrum emocji, takich jak zdziwienie, oburzenie, radość czy niedowierzanie. Znaczenie emotywów w dużym stopniu zależy od kontekstu ich użycia, co czyni je niezwykle złożonymi i trudnymi do jednoznacznej klasyfikacji. Wprowadzenie pojęcia *emotyw* przez zespół opolski miało na celu stworzenie hiperonimu (analogicznie jak w przypadku terminu „reprodukt”), który objąłby wszystkie gotowe, odtwarzalne struktury językowe służące wyrażaniu emocji nadawcy i jego reakcji na sytuację bądź komunikat. Ujęcie to zakłada oderwanie się od szczegółowych klasyfikacji strukturalnych na rzecz opisu funkcjonalnego, uwzględniającego również nacechowanie sytuacyjne – pozytywne bądź negatywne – oraz potencjalną wieloznaczność wypowiedzi. Dzięki temu emotywy analizowano nie tylko jako składniki językowego obrazu świata, lecz przede wszystkim w odniesieniu do ich znaczenia, funkcji w konkretnych sytuacjach komunikacyjnych, rodzaju wyrażanych emocji oraz możliwości odnalezienia rosyjskich odpowiedników, które występują w analogicznych kontekstach wraz z ich charakterystycznym brzmieniem (Chlebda i Milutina 2008: 45). Ujęcie to pozwoliło autorom idiomatykonu objąć pojęciem emotywu różnorodne konstrukcje reakcyjne, których przykłady – wraz z opisem leksykograficznym – zostały przedstawione poniżej w celu unaocznienia, jakie jednostki językowe obejmuje to pojęcie i jak są one rozumiane w ramach proponowanej koncepcji:

A nie mówiłem?	[potwierdzenie złych przeczuć wobec rozmówcy]
Co za różnica!	[obojętność wobec danej alternatywy]
Dzięki ci Boże!	[zadowolenie, ulga]
Kto by pomyślał!	[zdziwienie, zaskoczenie]
Nie bój nic!	[apel do rozmówcy, aby się nie przejmował]
Też mi coś!	[lekceważenie]
Tylko spróbuj!	[ostrzeżenie]

Warto zauważyć, iż w rosyjskim językoznawstwie funkcjonuje zbliżony termin odnoszący się do specyficznych reaktywnych jednostek językowych obecnych w żywej mowie – *коммуникема*. Pojęcie to zostało zaproponowane przez W. Ju. Melikjana w 2001 roku jako określenie syntaktycznych struktur niepełnych, które pełnią funkcje komunikacyjne w mowie potocznej. Autor ten, w *Słowniku emocjonalno-ekspresywnych zwrotów mowy żywej*, po raz pierwszy zebrał ponad 650 przykładów tego typu jednostek, wyrażających znaczenia twierdzenia, zaprzeczenia oraz oceny. Jak podkreśla Melikjan, *коммуникема* to «коммуникативная непредикативная единица синтаксиса, представляющая собой слово или сочетание слов, грамматически нечленимую, характеризующуюся наличием модусной позиции, нерасчленённо выражающую определённое непонятное смысловое содержание, не воспроизводящую структурных схем предложения и не являющуюся их регулярной реализацией, лексически непроницаемую и нераспространяемую, по особым правилам сочетающуюся с другими высказываниями в тексте и выполняющую в тексте реактивную, волюнтативную, эмоционально-оценочную, эстетическую и информативную функции» (Меликян 2001: 80). Jednostki te nie posiadają wartości predykatywnej ani nie są regularnymi realizacjami struktur zdaniowych, lecz występują jako reakcje – zarówno na wypowiedzi rozmówcy, jak i na fakty rzeczywistości – pełniąc różnorodne funkcje pragmatyczne. Z tego względu komunikem wpisuje się w ten sam funkcjonalny paradygmat co polski *emotyw*, definiowany jako spontaniczna, emocjonalnie nacechowana reakcja językowa w sytuacji bezpośredniej interakcji. Zgodnie z podejściem Melikjana, komunikemy można podzielić ze względu na ich funkcję, tj. twierdzenia i zaprzeczenia (np. *Ещё бы!*; *Как бы не так!*; *Безусловно*), emocjonalno-oceniające (np. *Вот это да!*; *Здрасьте!*), wolicjonalne, wyrażające imperatywne wezwania i rozkazy (np. *Айда!*; *Давай!*), kontaktowe (np. *Алло!*; *Внимание!*), etykietałne, podziękowania (*Не за что!*),

przeprosiny (*Пардон!*), powitania (*Здравствуйте!*), pytajne (np. *Ведь так?*; *Неужели?*; *Ну?*), tekstotwórcze, pełniące w tekście funkcję organizacyjno-kompozycyjną (np. *Итак*; *И потом...*), a także tzw. *коммуникативы*, które zakładają lub implikują reakcję słuchacza (*Ну ещё чего!* – jako ekspresywna odmowa). Przeciwstawione są im jednostki ekspresywne, które jedynie wyrażają emocje (np. *Вот это да!* jako reakcja na zaskoczenie lub zachwyt) i nie wymagają odpowiedzi ze strony odbiorcy (Бо 2015: 8).

W literaturze językoznawczej termin *комunikем* jest interpretowany w zróżnicowany sposób, w zależności od podejścia badawczego, np. G. N. Akimowa określa dane pojęcie jako «слова-предложения, самостоятельная функционально-структурно-семантическая разновидность простого предложения (односоставное), частицы и междометия, нечленимые, междометные предложения, устойчивые фразеологизированные обороты, междометия и словечки, особый разряд слов, слова-коммуникативы, релятивы, коммуникаты, неполные предложения и т.п.» (Акимова 2013: 195), akcentując formalno-strukturalny i semantyczny wymiar komunikemu jako jednostki składniowej.

Odmienne ujęcie prezentuje E. A. Chnykina: «коммуникема – явление живой разговорной речи, созданные как более экономные, краткие и эмоциональные формы выражения отношения, говорящего к объективному миру вокруг него» (Хныкина 2018: 111), podkreślając tym samym nacechowanie pragmatyczne terminu, osadzonego w żywej mowie potocznej.

Podobne podejście reprezentuje W. A. Krasnow, według którego *коммукемы* to «экономные и краткие формы выражения говорящего к объективному миру» (Краснов 2005: 5–9), stanowiące zwięzłe i uproszczone reakcje językowe, silnie osadzone w kontekście sytuacyjnym. *Комуникем* funkcjonuje zatem jako reakcja na różnego rodzaju fakty obiektywnej rzeczywistości i pełni w języku funkcje pragmatyczne, a jej wybór w mowie uzależniony jest od szeregu cech wspólnych wskazywanych przez badaczy: przynależność do języka mówionego, wysoka częstość użycia, intonacyjne oznaczenie oraz funkcja reakcji na mowę lub sytuację (Казачихина 2008: 115–119). Do ich podstawowych właściwości kategoryalnych zalicza się więc niepodzielność, nienominywność, sytuacyjność i ekspresyjność. Jednostki te cechują się również dużą dynamiką — ich skład jest systematycznie uzupełniany, a niektóre z danych jednostek uzyskują nowe znaczenia w wyniku neosemantyzacji. Pomimo tego, iż modele strukturalne pozostają stabilne, ich leksykalny skład ulega aktualizacji, co prowadzi do tworzenia wariantów semantycznych. Zmiany te wynikają z potrzeby precyzyjniejszego, a zarazem bardziej ekspresyjnego i oryginalnego przekazu. *Комуникем* stanowi zatem ważny komponent komunikacji, którego rola sprowadza się do

kompensowania braków systemu językowego, wzmacniania ekspresji wypowiedzi oraz skracania aktu mowy poprzez zastąpienie rozbudowanych struktur reprodukowalną i zwięzłą repliką (Гоголина 2014: 185).

Porównanie terminów *komunikem* i *emotywu* nie zostało przywołane przypadkowo, gdyż – jak już wcześniej wspomniano – *emotywu* funkcjonuje jako kategoria nadrzędna wobec licznych reakcji językowych, które w koncepcji Melikjana uznawane są za komunikemy, przy czym zyskują one dodatkowe klasyfikacje wynikające z ich struktury oraz funkcji komunikacyjnej. Ujęcie Melikjana, skoncentrowane na pragmatycznej roli jednostek tego typu, koresponduje z funkcjonalnym podejściem do języka, w którym istotne staje się nie tylko to, co się mówi, ale przede wszystkim *jak* i *dlatego* coś zostaje powiedziane.

W celu systematycznego uporządkowania zebranego materiału językowego opracowano poniżej autorską klasyfikację jednostek emotywnych. Klasyfikacja ta oparta została na kryteriach tematyczno-pragmatycznych, które odzwierciedlają intencję nadawcy, jego ocenę sytuacyjną, a także funkcję komunikacyjną wypowiedzi – taką jak wyrażenie sprzeciwu, aprobaty, dystansu, zniechęcenia czy uspokojenia. W przeciwieństwie do podziałów *stricte* afektywnych, kategorie te nie odnoszą się do emocji jako stanów wewnętrznych, lecz stanowią opis funkcjonalnych zachowań językowych w określonych sytuacjach komunikacyjnych. Opierając się na analizie materiału liczącego ponad 5000 polskich, rosyjskich i czeskich jednostek emotywnych, dokonano autorskiego podziału, który został wzbogacony o poniższe liczne przykłady jednostek o wyraźnie zarysowanej funkcji pragmatycznej:

KATEGORIA	PRZYKŁAD
1. Obojętność	<i>Mam to gdzieś; Мне плевать; Je mi to fuk</i>
2. Niewiedza	<i>A bo ja wiem; Бог знает; Nemám tušení</i>
3. Sprzeciw	<i>Nic z tych rzeczy; Никогда в жизни; V žádném případě</i>
4. Zadowolenie	<i>Wszystko gra; Как здорово!; To je paráda!</i>
5. Wycofanie	<i>Nie moja sprawa; Мне здесь делать нечего; To není moje věc</i>
6. Usprawiedliwienie	<i>Tak się tylko mówi; Это в переносном смысле; Jen se to tak říká</i>
7. Obawa	<i>Nawet nie mów; Только бы нет; Ani to neříkej</i>
8. Ulga	<i>Całe szczęście; Камень с души свалился; To bylo o fous</i>
9. Pewność	<i>Dam sobie rękę uciąć; Честное слово!; To je tutovka!</i>
10. Zgoda	<i>Jasna sprawa; Само собой; To si piš!</i>
11. Uspokojenie	<i>Spokojna głowa; Ничего страшного; Žádný stres</i>

12. Zaskoczenie	<i>Ładne buty; Bom тебе и на!; To je ale náhoda!</i>
13. Zdenerowanie	<i>Chyba kpisz; Это шутка, что ли?; To nemyslíš vážně?</i>
14. Rozczarowanie	<i>No i nic z tego; Всё напрасно; A to je smůla</i>
15. Ostrzeżenie	<i>Igrasz z ogniem; He тот подход; Ani to nezkoušej</i>

Przedstawiony podział jednostek emotywnych ma charakter porządkujący i nie wyczerpuje wszystkich możliwych sposobów klasyfikacji – ich pełne uchwycenie wymagałoby pogłębionej analizy znacznie szerszego materiału jednostek emotywnych. Jednak sama obserwacja zaproponowanych kategorii tematycznych oraz przyporządkowanych im reproduktów pozwala jednak dostrzec fundamentalny problem badawczy, związany z wieloznacznością danych wypowiedzi oraz ich zdolnością do funkcjonowania w różnych kontekstach komunikacyjnych, np. jednostka *Jaja sobie robisz* – może wyrażać zarówno zdenerwowanie, jak i zaskoczenie czy niedowierzanie, co znacząco utrudnia jednoznaczne przypisanie jej do konkretnej kategorii. Podobne trudności są powiązane z wariantywnością wielowyrazowców, która wynika z obecności komponentów stałych i zmiennych – w takiej sytuacji za podstawę analizy najczęściej przyjmuje się wariant najbardziej rozpowszechniony i typowy dla danej formy językowej. Uwzględnienie tego rodzaju zjawisk okazuje się szczególnie istotne w dalszych etapach pracy, zwłaszcza przy ustalaniu ekwiwalentów przekładowych dla wybranych polskich jednostek emotywnych, których interpretacja powinna opierać się również na analizie cech prozodycznych – takich jak intonacja, akcent, tempo czy siła artykulacji – ponieważ to one często pozwalają ograniczyć wieloznaczność wypowiedzi i ułatwiają odczytanie intencji nadawcy. Z tych właśnie względów, niniejsza praca koncentruje się na analizie jednostek nacechowanych emocjonalnie oraz ich opisie, który opiera się z jednej strony na podejściu stylistyczno-pragmatycznym (charakterystycznym dla tzw. szkoły opolskiej), a z drugiej – wzbogacony zostaje o elementy klasyfikacyjne wywodzące się z tradycji językoznawstwa rosyjskiego. Wraz z danymi jednostkami emotywnymi uwzględnione zostaną również wielowyrazowe jednostki fatyczne, które zostaną szczegółowo omówione w dalszej części pracy.

5.3. Fatyzmy

W kontekście analizowanych jednostek fatycznych szczególnie istotne staje się rozróżnienie pojęć *komunikacja* oraz *porozumiewanie się*, które – choć często traktowane równolegle w literaturze lingwistycznej – odnoszą się do odmiennych aspektów procesu wymiany językowej.

Komunikacja obejmuje działania skoncentrowane na intencjonalnym przekazie informacji i ich odbiorze oraz zakłada przede wszystkim przesyłanie konkretnych treści między nadawcą a odbiorcą.

Porozumiewanie się natomiast wykracza poza ten techniczny wymiar, akcentując interakcję jako przestrzeń budowania relacji międzyludzkich, dzielenia się doświadczeniami oraz wzmacniania więzi o charakterze społecznym i emocjonalnym (Стахнова 2013: 109–110).

Choć w ujęciach słownikowych terminy te bywają stosowane zamiennie – co potwierdzają definicje komunikacji jako „wymiany myśli, wiedzy, idei itd.” (Горелов 1990: 233–234) – zacieranie między nimi różnic może prowadzić do nieścisłości, zwłaszcza w analizie zjawisk językowych niezwiązanych bezpośrednio z przekazywaniem informacji. Przykładem takiego zjawiska jest funkcja fatyczna, która nie tyle realizuje się poprzez przesyłanie treści, ile poprzez sam akt podtrzymywania kontaktu, ujawniając swoje rzeczywiste znaczenie i funkcję właśnie w obszarze porozumiewania się rozumianego jako interakcja o charakterze relacyjnym.

Początek refleksji nad fatycznym aspektem komunikacji wiąże się z pracami brytyjskiego antropologa Bronisława Malinowskiego, który jako pierwszy wprowadził pojęcie *phatic communion*, zwracając uwagę na taki sposób użycia języka, który nie służy przekazywaniu informacji o charakterze poznawczym, lecz pełni funkcję społeczną – umożliwia nawiązanie i podtrzymanie więzi między rozmówcami (Malinowski 1972: 146). W tego rodzaju wypowiedziach aspekt znaczeniowy zostaje zredukowany, a czasami wręcz całkowicie zanika, ponieważ nie służą one przekazywaniu informacji, lecz mają na celu podtrzymanie kontaktu między rozmówcami. Jak zauważa Malinowski, pełnią one przede wszystkim funkcję rytualną i relacyjną, a tym samym stają się ważnym elementem codziennego życia społecznego. Do myśli Malinowskiego nawiązał Roman Jakobson, który, rozwijając wcześniejszy model Karla Bùhlera, włączył funkcję fatyczną do swojej sześcioczęłowej typologii funkcji językowych. Bühler, konstruując swój model w oparciu o trójkąt komunikacyjny, wyróżnił trzy podstawowe funkcje języka:

- *ekspresywną*, zorientowaną na nadawcę;
- *impresywną*, ukierunkowaną na odbiorcę oraz
- *przedstawieniową*, odnoszącą się do rzeczywistości pozajęzykowej (Bühler 2004)

Jakobson rozszerzył daną koncepcję, wskazując na sześć podstawowych funkcji języka, wśród których – obok referencyjnej, emotywniej, konatywnej, poetyckiej i metajęzykowej – znalazła się również funkcja fatyczna, opisana jako *nastawienie na kontakt* (Jakobson 1960:

438). W przeciwieństwie do Malinowskiego, który akcentował społeczny wymiar wypowiedzi fatycznych, Jakobson skoncentrował się na ich aspekcie technicznym, traktując je jako środki służące ustanowieniu, podtrzymaniu lub wznowieniu kontaktu między uczestnikami aktu komunikacji. Funkcja ta realizuje się zatem poprzez komunikaty umożliwiające sprawdzenie działania kanału komunikacyjnego i potwierdzenie obecności odbiorcy – przykładem mogą być wypowiedzi typu „*Halo, słyszysz mnie?*” czy „*Jesteś tam jeszcze?*”, których celem nie jest przekazanie informacji, lecz utrzymanie interakcyjnej ciągłości. Choć przez długi czas funkcja fatyczna była marginalizowana w badaniach nad językiem, współczesna literatura coraz częściej interpretuje ją nie tylko jako narzędzie utrzymywania technicznego kontaktu, ale również jako formę aktywności językowej mającej wymiar społeczny, emocjonalny i relacyjny.

Rozumienie funkcji fatycznej w polskim językoznawstwie najlepiej oddają definicje zawarte w *Encyklopedii wiedzy o języku polskim* oraz *Encyklopedii językoznawstwa ogólnego*, zgodnie z którymi są to „środki służące do ustalania i podtrzymywania kontaktu językowego między nadawcą a odbiorcą” (Kurdyła 2015: 61). Do tej kategorii zalicza się zarówno formuły inicjujące kontakt (np. *Dzień dobry; Witaj; Szanowny Panie*), jak i środki podtrzymujące dialog (*aha; no jasne; oczywiście; tak, tak*) oraz wypowiedzi kończące rozmowę (*no to tyle; cześć, z wyrazami szacunku*).

Kategoria ta była rozwijana w polskiej literaturze m.in. przez Krystynę Pisarkową, Jadwigę Puzyninę, Romualda Pietkowską, Bożenę Witosz, Aleksego Awdiejewa, Grażynę Habrajską oraz – częściowo – Jolantę Tambor. Ta ostatnia zauważa, że funkcję fatyczną pełnią nie tylko pojedyncze formuły, lecz całe akty mowy, ponieważ „rozmówcy mówią tylko dlatego, że wypada, trzeba mówić”, a treść ich wypowiedzi bywa drugorzędna wobec potrzeby podtrzymania kontaktu (Tambor 1991: 181).

Na podobną intencjonalność wskazuje Renata Grzegorzyczkowska, pisząc, że funkcja fatyczna „najczęściej polega na swoistej intencji wypowiedzi (często ukrytej), której temat może być zupełnie inny” (Grzegorzyczkowska 1991: 13). Choć nie wyodrębnia jej jako odrębnej funkcji językowej, traktuje funkcję fatyczną jako jedną z dwunastu funkcji wypowiedzi, których celem jest podtrzymywanie więzi między uczestnikami interakcji. Ujęcie to rozwija Monika Otulak-Komenda, wskazując, że fatyczność obejmuje także zachowania para- i pozawerbalne, a jej istotnym aspektem jest kontrola przebiegu rozmowy i kształtowanie przyjaznej atmosfery dla rozmówców. Funkcja ta przejawia się na różnych poziomach – od środków językowych, przez teksty, po sam akt mowy – i pełni rolę organizującą interakcję społeczną (Wojtczuk 2010: 87). Z kolei Urszula Żydek-Bednarczuk podkreśla, że teksty o funkcji fatycznej cechuje formalna schematyzacja i realizacja specyficznych reguł

komunikacyjnych (Żydek-Bednarczuk 2005: 200), natomiast Drabik-Frączek i Załazińska akcentują społeczny wymiar fatyczności, uznając komunikację za praktykę tworzenia relacji i więzi międzyludzkich (Drabik-Frączek i Załazińska 2014: 40). Nieco odmienną perspektywę prezentuje Antoni Furdal, który traktuje funkcję fatyczną jako odmianę funkcji komunikatywnej, a nie jako odrębną kategorię: „jeżeli mówi się do kogoś banały i ogólniki, komunikuje mu się tym samym treści ogólne, w których ładunek informacyjny jest znikomy, ale jest. A to, czy celem odezwań jest podtrzymanie kontaktu, czy odwrócenie czyjejś uwagi, czy nawet oszukanie kogoś, to są sprawy, które wykraczają poza zakres funkcji języka” (Furdal 1977: 46). Jego stanowisko wskazuje na trudność jednoznacznego wyodrębnienia funkcji fatycznej i poddaje w wątpliwość sens jej odrębnego traktowania w systemie funkcji językowych.

Wśród polskich opracowań poświęconych wyrażeniom o funkcji fatycznej brakowało dotąd pracy, która w sposób systematyczny i konfrontatywny podejmowałaby zagadnienie fatyzmów w języku polskim, czeskim i rosyjskim. Lukę tę wypełnia monografia Andrzeja Charciarka, będąca pierwszą próbą szczegółowej analizy tego typu jednostek z perspektywy pragmatyczno-tekstologicznej. Autor, bazując na materiale empirycznym liczącym 15 000 fragmentów dialogów (po 5000 z każdego języka) (Charciarek 2010: 17), koncentruje się na tzw. operatorach metatekstowych o funkcji fatycznej, a więc takich jednostkach, które regulują kontakt werbalny między uczestnikami rozmowy, inicjując go, podtrzymując lub kontrolując jego przebieg. Analizie poddano przede wszystkim struktury typu *wiesz; rozumiesz; słuchaj; popatrz; no nie?*, rozpatrywane pod kątem ich funkcji semantycznych, pozycji w wypowiedzi oraz stopnia desemantyzacji. Materiał pochodzi głównie ze spontanicznych rozmów w codziennych sytuacjach komunikacyjnych, co nadaje analizie autentyczności i pozwala uchwycić różnice pragmatyczne w porównywanych językach oraz wskazać braki w istniejących opracowaniach leksykografii przekładowej. Warto odnotowania są także indeksy operatorów fatycznych, liczące łącznie 457 jednostek: 168 w języku polskim, 129 czeskich i 160 rosyjskich. Całość opracowania wpisuje się w nurt badań nad językiem mówionym i zarazem wskazuje na istotne luki w tradycyjnych opisach słownikowych, w których jednostki fatyczne bywają systematycznie pomijane.

W nurcie współczesnych rosyjskich badań nad funkcją fatyczną szczególnie istotne okazują się ujęcia T. G. Winokur oraz L. A. Aznabajewy, które sytuują ten aspekt komunikacji poza tradycyjnie pojmowanym przekazem informacji. Tatiana Winokur wskazuje na istnienie dwóch przeciwstawnych biegunów w systemie funkcji językowych – informacyjnego i nieinformatywnego – z których ten drugi reprezentowany jest właśnie przez funkcję fatyczną,

pełniącą kluczową rolę w strukturze wypowiedzi i kształtowaniu relacji interpersonalnych. Badaczka zauważa, że opozycja funkcja informacyjna – funkcja fatyczna stanowi podstawę dominującej funkcji mowy, która nie tyle służy przekazowi, ile realizuje intencję nawiązania kontaktu i uczestnictwa w interakcji: «таких особенностей, связанных с основной (доминирующей) функцией, обслуживающей процесс коммуникативного взаимодействия автора и адресата, а именно — с интенцией вступить в речевое общение, является оппозиция «информативная — фатическая речь» [...]» (Винокур 1993: 135).

W podobnym duchu wypowiada się L. A. Aznabajewa, która rozróżnia aspekty kognitywne i indeksykalne w ludzkich zachowaniach werbalnych, przypisując tym drugim – utożsamianym z funkcją fatyczną – rolę podtrzymywania kontaktu i orientowania komunikacji na jej relacyjny wymiar (Азнабаева 1989: 36). Zgodnie z tymi założeniami mowa fatyczna nie jest zjawiskiem pobocznym ani pomocniczym, lecz integralnym elementem procesu porozumiewania się. Jej treść nie musi zawierać wartości informacyjnej, gdyż – jak zauważa Winoкур – «фатическая речь расширяет рамки объекта, располагая внутри него конативную функцию элементарного содержания (вступление в контакт, его поддержка и проверка)» (Винокур 1993: 135). Tego rodzaju interpretacje podkreślają konieczność wyraźnego rozróżnienia komunikacji i porozumiewania się, gdzie opozycja przekaz – brak przekazu przybiera wymiar funkcjonalny, a nie treściowy. W analizie funkcji fatycznej punktem wyjścia jest zatem założenie, że akt mowy zawiera dwa aspekty: właściwy komunikacyjny (informacyjny) oraz metakomunikacyjny, związany z regulacją przebiegu interakcji. Ta regulacja dotyczy m.in. nawiązania kontaktu werbalnego, utrzymania uwagi rozmówcy oraz zakończenia dialogu, przy czym proporcja między warstwą komunikacyjną a metakomunikacyjną zależy od konkretnych warunków sytuacyjnych – takich jak cechy uczestników, ich role społeczne, relacje, intencje czy kontekst tematyczny.

W językoznawstwie czeskim zagadnienie funkcji fatycznej nie znalazło jak dotąd równie szerokiego opracowania jak funkcje ekspresywne, poznawcze czy metajęzykowe, jednak pojawia się w pracach inspirowanych klasycznym modelem komunikacji Romana Jakobsona oraz w badaniach nad rozmową jako specyficzną formą interakcji językowej.

Jak zauważa J. Hoffmannová, choć funkcja fatyczna długo pozostawała na marginesie zainteresowań, stanowi ona klucz do zrozumienia społecznych i kulturowych uwarunkowań konwersacji, uznawanej za jej główny przejaw (Hoffmannová 1996: 191). Czescy badacze, zwłaszcza kontynuatorzy tradycji Praskiej Szkoły Lingwistycznej, postrzegają konwersację nie tylko jako techniczny akt utrzymywania kanału komunikacyjnego, ale przede wszystkim jako narzędzie podtrzymywania więzi, rytualizacji kontaktu oraz budowania wspólnoty.

O funkcji tej pisał W. Mathesius, określając ją jako „rozmowę dla rozmowy” czy „dla przyjemności z rozmowy” i podkreślając jej wymiar społeczny (Mathesius 1982: 393), natomiast J. Mukařovský dostrzegał w niej również funkcję estetyczną, zwracając uwagę na wartość formy i rytuału komunikacyjnego (Mukařovský 1948: 137).

Na konwersację jako odrębny styl funkcjonalny zwrócili uwagę przedstawiciele młodszej generacji Praskiej Szkoły Lingwistycznej, podejmując próbę określenia jej statusu wobec innych odmian języka mówionego. M. Jelínek analizował relację między stylem konwersacyjnym a prostym stylem informacyjnym (Jelínek 1966: 107), natomiast J. Hrbáček wskazywał na wtórność funkcji informacyjnej w konwersacji oraz jej estetyczny potencjał, zbliżając się w tym do ujęcia Mukařovskiego (Hrbáček 1983: 111–114). Współczesna refleksja, reprezentowana przez Hoffmannovą, poszerza to ujęcie o analizę całego spektrum zjawisk językowych i społecznych – od banalnych *small talków* i rytualnych powitań, przez rozmowy w kolejce, tramwaju, przy kawie, aż po konwencjonalne *drby*, czyli codzienne plotki i opowieści, nierzadko pełne ekspresji i emocjonalnego zaangażowania. Jak pisze autorka, „Celá sféra fatické funkce se svou kontaktovou orientací vymyká praktičnosti a instrumentálnosti většiny komunikačních událostí” (Hoffmannová 1996: 204), co wskazuje na jej specyficzny charakter – oderwanie od celów informacyjnych na rzecz budowania relacji i strukturyzowania sytuacji społecznych. Czeska refleksja nad fatycznością nie ogranicza się zatem do analizy środków językowych, lecz obejmuje również kwestie stylu, etykiety, estetyki wypowiedzi oraz jej zakorzenienia w rytuałach codzienności, tworząc obraz funkcji fatycznej jako zjawiska z pogranicza języka, kultury i społecznego rytuału.

W ramach myśli opolskiej pojęcie *fatyzmu* (analogicznie do *emotywu*) funkcjonuje jako hiperonim obejmujący wielowyrazowe jednostki językowe, których podstawową funkcją jest podtrzymywanie kontaktu w rozmowie bądź otwarcie kanału komunikacyjnego. Należy mocno podkreślić, iż w odróżnieniu od tradycyjnych ujęć, w których fatyzmy zawierają także formuły grzecznościowe, powitania czy pożegnania, zespół opolski oddziela te struktury i klasyfikuje je jako adresatywy, wyłączając je tym samym z zakresu badanych jednostek. W czwartym tomie opolskiego *Idiomatykonu*, w rozdziale zatytułowanym *Rozmowy nasze codzienne*, fatyzmy zostają potraktowane jako powtarzalne elementy językowe, które – podobnie jak emotywy – nie są klasyfikowane według ściśle określonych kategorii formalnych, lecz funkcjonują jako ustabilizowane środki służące organizacji wypowiedzi i utrzymaniu płynności rozmowy. Choć autorzy nie dokonują pogłębionej analizy pragmatycznej tychże jednostek, ich opis skupia się na wskazaniu rosyjskich ekwiwalentów przekładowych – łącznie zidentyfikowano 164 wyrażenia tego typu wraz z odpowiednikami rosyjskimi. Ze względu na

stosunkowo niski poziom wieloznaczności poszczególnych fatyzmów zrezygnowano z przedstawienia typowych kontekstów ich użycia. Choć decyzja ta znajduje uzasadnienie w stabilnym funkcjonowaniu tych struktur w języku, to z perspektywy niniejszej pracy uwzględnienie opisu sytuacyjnego wydaje się zabiegiem pożądanym, gdyż pozwala lepiej uchwycić ich funkcję w realnych aktach komunikacyjnych. Na podstawie obserwacji zgromadzonego materiału językowego możliwe staje się wyodrębnienie trzech ogólnych grup jednostek fatycznych:

- 1) otwierające kanał komunikacyjny (np. *moim zdaniem; wydaje mi się, że; rzecz w tym, że*);
- 2) podtrzymujące jego drożność (*a na dodatek; odnośnie do tego; a najważniejsze*) oraz
- 3) reagujące w sposób zachęcający do dalszego udziału w rozmowie (*kto by pomyślał; co pan mówi?; nic mi to nie mówi*).

Warto podkreślić, że większość z nich ma charakter składniowo niepełny – wymagają one dopełnienia, które może pojawić się przed lub po danym zwrocie, co czyni je typowymi wtrąceniami. Takie struktury, choć często pomijane w opisach systemu języka, są kluczowe dla płynnej i naturalnej komunikacji dwujęzycznej. Ich brak, szczególnie w języku obcym, może skutkować niezręcznymi przerwami w rozmowie, które pojawiają się w momentach przejściowych, na przykład wtedy, gdy nadawca potrzebuje czasu na zebranie myśli. W takich sytuacjach znajomość jednostek fatycznych pozwala nie tylko podtrzymać kontakt, ale także sprawia wrażenie aktywnego uczestnictwa w dialogu, co ma istotne znaczenie zarówno w kontekście językowej płynności, jak i kompetencji interakcyjnej. Pomimo uproszczonej formy prezentacji, braku klasyfikacji tematycznej oraz ograniczenia opisu do odpowiedników rosyjskich, podejście opolskie stanowi cenne źródło dokumentacyjne, które stało się punktem wyjścia do zgromadzenia szerszego materiału badawczego oraz opracowania autorskiej klasyfikacji, opartej na analizie około 1500 jednostek fatycznych pochodzących z języka polskiego, rosyjskiego i czeskiego:

KATEGORIA	PRZYKŁAD
1. Otwierające	<i>Moim zdaniem; Дело в том, что; Pokud si dobře pamatuju</i>
2. Podtrzymujące	<i>Swoją drogą; Более того; Kromě toho</i>
3. Reagujące	<i>No coś ty; Вот это да!; Fakt jo?</i>
4. Aktywizujące	<i>Jestem ciekaw; Ну говори; Jak to dopadlo?</i>
5. Porządkujące	<i>Z jednej strony; Другими словами; Stručně řečeno</i>
6. Przerywnikowe	<i>No wiesz; Ну да; Jak to říct</i>
7. Kończące	<i>No i tym podobne; На этом всё; Tím to hasne</i>

Zebrane i zaprezentowane powyżej przykłady jednostek fatycznych ukazują nie tylko ich wewnętrzne zróżnicowanie, lecz także szereg trudności związanych z ich klasyfikacją, delimitacją czy interpretacją znaczeniową. Pomimo przyjęcia funkcji komunikacyjnej jako podstawowego kryterium podziału (otwierającej, podtrzymującej, reagującej, porządkującej czy kończącej), wiele jednostek wykazuje wielofunkcyjność i podatność na przesunięcia znaczeniowe. W zależności od otoczenia kontekstowego bądź pozycji w tekście, te same ciągi werbalne mogą być interpretowane jako przerywnikowe, aktywizujące lub porządkujące, a w skrajnych przypadkach – mogą pełnić wszystkie te role równocześnie. Cecha ta nie tylko utrudnia jednoznaczne przypisanie jednostek do określonych kategorii, lecz także podważa sztywność klasyfikacyjną jako taką, wskazując na potrzebę podejścia dynamicznego i kontekstowego. Dodatkową trudność stanowi zależność form fatycznych – zwłaszcza reagujących – od cech prozodycznych, takich jak intonacja, akcent czy tempo, które nierzadko decydują o ich funkcji i często zbliżają je do emotywów pod względem formy i znaczenia. W praktyce oznacza to, że te same wypowiedzi mogą być jednocześnie nośnikiem relacyjnej funkcji fatycznej oraz emocjonalnej funkcji ekspresywnej, co komplikuje ich rozróżnienie i wymaga każdorazowego odniesienia do ramki pragmatycznej. Jednym z głównych problemów badawczych pozostaje silny stopień wariantowości wielowyrazowych jednostek fatycznych, która sprawia, że mimo pozornej stabilności i rozpoznawalności, fatyzmy okazują się kategorią wyjątkowo dynamiczną, rozproszoną i silnie zależną od kontekstu.

Przedstawione w niniejszym rozdziale wielowyrazowe jednostki fatyczne i ematywne stanowią szczególne struktury językowe, których silne zakorzenienie w potocznej komunikacji nie ulega wątpliwości, a zarazem są trudne do jednoznacznego uchwycenia w ramach przyjętych klasyfikacji. Obie grupy wykazują szereg zbliżonych cech, takich jak wysoka frekwencja w języku mówionym, polifunkcyjność, zależność od kontekstu oraz zmienność formalna, która utrudnia ich kodyfikację słownikową. To właśnie te właściwości sprawiają, że

jednostki te – mimo iż obecne w codziennym użyciu – pozostają wciąż niedostatecznie opisane, zwłaszcza w kontekście leksykografii jedno- i dwujęzycznej, która rzadko uwzględnia ich rzeczywiste funkcjonowanie w komunikacji. Z tego względu dalsza część pracy poświęcona będzie metodom wyodrębniania fatyzmów i emotywów, co pozwoli uchwycić ich pragmatyczne i semantyczne właściwości oraz zbliżyć się do pełniejszego opisu reproduktów funkcjonujących w języku polskim, rosyjskim i czeskim.

ROZDZIAŁ VI

Metody wyodrębniania reproductów fatycznych i emotywnych

Wyodrębnianie wielowyrazowych jednostek fatycznych i emotywnych pozostaje kluczowym wyzwaniem w lingwistyce stosowanej, ponieważ jak dotąd nie opracowano w pełni efektywnych metod automatycznych. Skomputeryzowane metody wykorzystujące selektory tekstów lub znaków okazują się skuteczne w poszukiwaniu kolokacji, frazeologizmów czy skrzydlatych słów, jednak nie znajdują zastosowania w przypadku analizowanych przez nas reproductów języka mówionego. Niemniej jednak istnieje szereg procedur wspierających, które mogą znacząco usprawnić proces poszukiwania tego rodzaju jednostek. O ile granice pojedynczych wyrazów są z natury jasno określone przez ich formę, to w przypadku reproductów wielowyrazowych kluczowym problemem jest delimitacja, czyli precyzyjne wyznaczenie ich początku i końca jako spójnych i odrębnych znaczeniowo konstrukcji. Proces ekscerpacji tego typu jednostek napotyka szereg trudności związanych z ich identyfikacją, do których można zaliczyć między innymi:

- 1) **Znaczną zmienność formalną reproductów**, uwzględniającą ich wariantywność, elastyczność strukturalną oraz możliwość modyfikacji składniowej i gramatycznej, w tym wymiennność oraz wtrącanie nowych komponentów pomiędzy elementy stałe, np.

CO ZA PYTANIE / CO ZA PROSTACKIE PYTANIE / I CO TO W OGÓLE ZA PYTANIE		
— <i>Jak śmiesz! Co za pytanie.</i> <i>Zostawiłeś mnie samą z tą uroczą dwójką!</i>	- <i>No wiesz... Co za prostackie pytanie! Kosmici nas mogą unicestwić, a unici nas poprowadzą ku szczęśliwym horyzontom!</i>	<i>Nie bądź żalosny... Bądź facetem, Andrzej! I co to w ogóle za pytanie? Po tym wszystkim, co zrobiłeś dziecku, masz jeszcze czelność pytać?</i>

ПОНЯТИЯ НЕ ИМЕЮ / НЕ ИМЕЮ НИ МАЛЕЙШЕГО ПОНЯТИЯ / НЕ ИМЕЮ АБСОЛЮТНО НИКАКОГО ПОНЯТИЯ		
— <i>Я спрашиваю, что я должен делать на работе? —</i> <i>Понятия не имею, —</i> <i>ответил он.</i>	<i>Где была моя гордость эти полтора года — не имею ни малейшего понятия.</i>	<i>Не имею абсолютно никакого понятия, кто это написал, но я горжусь этой женщиной!"</i>

JE MI TO JEDNO / JE MI TO ÚPLNĚ JEDNO / UPŘÍMNĚ JE MI ÚPLNĚ JEDNO		
<i>Mně se nezdá. Ale i kdyby to bylo pravda, je mi to jedno. Co jsem viděl až dosud, stačilo by mi k spokojenosti víc než dost.</i>	<i>Já: Nebylo by lepší, kdybys mi předem nic neříkala a udělala to podle svého? Víš přece, že je mi to úplně jedno.</i>	<i>Upřímně je mi úplně jedno, kdo v #gopwars hází špínu na koho. Nejsem na ničí straně.</i>

- 2) **Zdolność reproduktów do występowania w formie rozproszonej**, gdzie ich komponenty nie są skupione w jednym miejscu, lecz rozrzucone w obrębie jednego zdania lub sąsiadujących zdań, co wynika z ich strukturalnej przenikalności.

COŚ Z <TYM> FANTEM ZROBIĆ
<i>Nie ulega wątpliwości, że z fantem w postaci zadłużonych szpitali trzeba coś zrobić i potrzebne są nowatorskie pomysły.</i>

ОТЛОЖИТЬ В ДОЛГИЙ ЯЩИК
<i>Нацистское нашествие заставило Сталина, в самом деле вообразившего себя гением, а в действительности только что опозорившегося с его «скрепленным кровью» союзом с Гитлером, отложить <u>планы «мировой революции»</u> в <u>долгий ящик</u>.</i>

ZTRÁCET <PEVNOU> PŮDU POD NOHAMA
<i>Jakmile pocítil, že pevná půda, <u>na které doposud stál, se mu začíná ztrácet pod nohama</u>; když vycítil, že hlavní důvod vlastní existence, jímž se doposud ospravedlňoval, ztrácí smysl.</i>

- 3) **Wieloznaczność reproduktów**, nieostrość granic między ich różnymi znaczeniami oraz podatność na semantyczne zmiany (neosemantyzacja). (Chlebda 2010b: 19).

ZAWSZE COŚ	
<i>Chętnie obciążamy dziecko odpowiedzialnością, czepiając się byle pretekstu. A to było nie dość grzeczne, a to nie odrobiło lekcji. Zawsze coś! (<u>stosowane jako reakcja na regularne unikanie wykonania określonego obowiązku przez rozmówcę</u>).</i>	<i>Tak się głodzisz i tak mało schudłaś? No to rzeczywiście. Ja też strasznie mało tracę na wadze, ale spodnie są już luźniejsze. Może nie tak, jakbym chciała, ale zawsze coś. (<u>używane dla określenia drobnej korzyści, która może pozytywnie wpłynąć na przebieg sytuacji</u>).</i>

КАК СКАЖЕШЬ	
— А теперь все сначала. И без военных песен. — Как скажешь, — легко согласился Дубинюк. — Твое слово — закон. <u>(używane w celu akceptacji zdania rozmówcy).</u>	— Ты же сама жаловалась, что их собака на твои цветы гадит, — удивился муж. — Пусть гадит. — Ладно, как скажешь. Только потом не жалуйся, что они тебе надоели. <u>(pozorna zgoda z elementem ironii lub sceptyzmu).</u>

JE TO NA TOBĚ	
Ano, ale..., Žádné ale, synku. Je pravda, že jsou různá pojetí cti - ovšem člověk musí žít jenom podle jednoho z nich. Dělej, co chceš. Je to na tobě. <u>(używane dla wyrażenia neutralnej akceptacji decyzji rozmówcy).</u>	Tým Macu musel vyrůst a to musíš ty taky, “odtušil Jobs. „Chci, aby ses vrátil ,ale jestli nechceš, je to na tobě. Nejsi zase tak důležitý, jak si o sobě myslíš. <u>(używane dla wyrażenia ukrytego oczekiwania wobec rozmówcy i wywarcia nacisku emocjonalnego).</u>

W tym miejscu pojawia się zasadnicze pytanie, czy możliwe jest opracowanie skutecznych metod pozwalających na wyodrębnienie wielowyrazowych jednostek fatycznych i emotywnych z wysokim stopniem prawdopodobieństwa trafności. Kluczowym warunkiem wstępnej selekcji tekstów oraz identyfikacji interesujących nas jednostek jest spełnienie określonych kryteriów analitycznych, umożliwiających precyzyjne rozpoznanie potencjalnych reпродуктów, czyli tzw. *gotowości intuicyjnej*.

Zastosowanie intuicji badacza jako narzędzia w procesie analizy wybranego zagadnienia językowego od lat stanowi przedmiot dyskusji, wzbudzając zarówno uznanie ze strony jej zwolenników, którzy podkreślają jej wartość odkrywania niestandardowych rozwiązań i naturalnej umiejętności weryfikacji założonych hipotez, jak i krytykę ze strony przeciwników, wskazujących na subiektywizm i brak wiarygodnej weryfikacji uzyskanych wyników. Należy jednak podkreślić, że to właśnie intuicja odgrywa fundamentalną rolę nie tylko w procesach przyswajania języka i komunikacji werbalnej, ale także w tworzeniu wypowiedzi, wspierając wybór odpowiednich struktur gramatycznych i leksykalnych oraz nieświadomym kodowaniu regularności występujących w otoczeniu w pamięci niejawnej (Piegzik 2022a: 137). W językoznawstwie wyróżnia się głównie dwie koncepcje intuicji, z których pierwsza, inspirowana Noamem Chomskym, traktuje intuicję jako zinternalizowaną wiedzę językową, czyli kompetencję umożliwiającą nieświadome ocenianie gramatycznej poprawności zdań dzięki znajomości uniwersalnej gramatyki. Druga, wywodząca się z tradycji

francuskiej, postrzega intuicję jako pierwotne struktury mentalne związane z przedjęzykowością, przygotowujące umysł do przyswojenia języka i poznania. Koncepcja ta odwołuje się do idei Henri Bergsona, podkreślającego intuicję jako kluczowy mechanizm ewolucyjny i poznawczy. Intuicyjna wiedza językowa umożliwia szybkie ocenianie poprawności i sensowności wypowiedzi oraz kompensowanie braków językowych poprzez domyślanie się znaczeń w kontekście. Intuicja językowa, rozwijająca się spontanicznie poprzez praktykę językową, pełni głównie funkcję prognozującą, kontrolną i integracyjną, umożliwiając szybkie przetwarzanie informacji oraz rozumienie niestandardowych form językowych, takich jak metafory czy nietypowe konstrukcje. Mechanizmy intuicji, oparte na nieświadomych procesach, odgrywają istotną rolę zarówno w codziennej komunikacji, jak i w interpretacji tekstów artystycznych, wspierając rozpoznawanie ukrytych znaczeń i estetycznej wartości wypowiedzi (Быра 2011: 8). Warto w tym miejscu podkreślić dynamiczny charakter intuicji, która rozwija się w wyniku nieświadomego uogólniania doświadczeń językowych podczas regularnej praktyki i nie stanowi cechy stałej ani wrodzonej, lecz podlega ciągłemu rozwojowi (Чибухашвили 2009: 63–69). Niektórzy badacze podkreślają również dualistyczny charakter intuicji językowej, która obejmuje zarówno nieświadome rejestrowanie jednostek językowych, jak i świadome operowanie wiedzą o normach językowych oraz zasadach doboru odpowiednich środków wyrazu (Баранов, Ладыженская, Львов 1990: 59).

Kluczowym aspektem intuicji pozostaje tzw. *wyczucie języka* (*language feel*), które rodzimi użytkownicy rozwijają naturalnie podczas przyswajania pierwszego języka (Lewandowska-Tomaszczyk 2010: 68), co sprawia, że rozwój tej umiejętności w przypadku nauki języka obcego jest znacznie trudniejszy. Warto w tym miejscu wspomnieć o dwóch rodzajach intuicji w kontekście języka rodzimego: intuicji użytkowników, czyli naturalnej zdolności użytkowników języka, oraz intuicji badawczej, wykorzystywanej przez lingwistów w tradycyjnych badaniach, która obecnie ustępuje miejsca bardziej obiektywnym metodom (Sgall 2010: 86). Podjęte badania w zakresie glottodydaktyki wskazują jednak na istotną rolę intuicji w procesie przyswajania języka obcego, szczególnie w obszarze nabywania wiedzy implicytnej, co wspiera płynność i naturalność użycia języka. Wyniki testów sugerują, że intuicja jest najbardziej efektywna w przypadku form regularnych i często używanych, podczas gdy jej skuteczność maleje w obliczu złożonych struktur gramatycznych (Piegzik 2020: 233). Należy również podkreślić, że podstawą intuicji jest tzw. *pamięć semantyczna*, w której pobudzenie węzłów, rozumianych jako określone pojęcia, wywołuje skojarzeniowe połączenia z innymi węzłami, czyli pozostałymi terminami (Sarzyńska 2009: 135–136). To wszystko umożliwia językoznawcy szybkie uchwycenie wzorców, analogii i relacji w danych

językowych, co jest szczególnie istotne w sytuacjach deficytu lub nadmiaru informacji, a jednocześnie dzięki mechanizmowi precyzyjnego identyfikowania częstych połączeń wyrazowych, utrwalonych w pamięci poprzez ich wielokrotne rejestrowanie, pozwala na generowanie hipotez i przeczuć, które następnie mogą być weryfikowane za pomocą nowoczesnych metod analitycznych (Piegzik 2022b: 181–192). W kontekście badań nad intuicją językową nierzadko używane jest również pojęcie introspekcji, rozumiane jako proces samopoznania, w którym językoznawca odwołuje się do własnych wewnętrznych doświadczeń w celu analizy określonych zjawisk i obejmuje między innymi takie elementy, jak:

- 1) Subiektywność – wykorzystanie indywidualnych odczuć w celu oceny poprawności gramatycznej czy semantycznej zdań;
- 2) Analiza danych językowych – konstruowanie przykładowych modeli zdaniowych do weryfikacji ich poprawności;
- 3) Komponent kognitywny – wewnętrzne monitorowanie procesów myślowych i językowych podczas badania relacji między językiem a myśleniem itp. (Тимофеева 2015: 33–55).

Zasadniczą rolę odgrywa tutaj również logika, która jest ściśle powiązana z intuicją jako wzajemnie uzupełniające się aspekty procesu poznawczego, w którym logika odgrywa fundamentalną rolę w analizie, uwzględniającej porównywanie jednostek, wybór strategii oraz korektę popełnionych błędów. Intuicja wprowadza natomiast element syntezy, szczególnie istotny w sytuacjach, gdy analiza logiczna napotyka ograniczenia wynikające z braku pełnych danych. Choć intuicyjne rozumienie bywa niezbędne w takich przypadkach, wymaga ono jednocześnie ostrożności i weryfikacji, aby zminimalizować ryzyko błędnych wniosków. Współdziałanie logiki i intuicji umożliwia zatem bardziej kompleksowe i skuteczne rozwiązywanie problemów poznawczych.

Pojęcia intuicji językowej, świadomości językowej oraz kompetencji językowej są często używane wymiennie, co znajduje odzwierciedlenie w literaturze przedmiotu. Ich wzajemne powiązanie można zilustrować na przykładzie definicji przedstawionej przez Stanisława Gajdę, który precyzuje istotę tych zjawisk w kontekście procesów językowych: „Można zatem świadomość językową traktować i jako odbicie rzeczywistości praktyki językowej, i jako swoisty mechanizm programujący i regulujący zachowania językowe, czyli jako naturalną wewnętrzną normę językową” (Gajda 1987: 586).

Z kolei Porayski-Pomsta postrzega świadomość językową jako pojęcie nadrzędne, odnoszące się do postawy wobec języka i zależne od kompetencji językowej, która z kolei wpływa na efektywność wykonania językowego. Kompetencja pozwala użytkownikowi oceniać aspekty językowe, takie jak poprawność, sensowność czy synonimiczność wypowiedzi

(Szymańska-Galińska 2023: 173–174). Proces ten rozwija się od intuicyjnego poczucia językowego przez świadomość semiotyczną aż po metajęzykową, umożliwiając głębsze rozumienie języka, analizę jego struktur oraz efektywne przekazywanie wiedzy, co ma kluczowe znaczenie w edukacji i kształtowaniu kompetencji lingwodydaktycznej (Śniatkowski 2011: 499). Struktura świadomości językowej składa się ze sfery *kognitywnej* (uwzględniającej wiedzę o języku i nabytą kompetencję językową), *pragmatycznej* (dotyczącej zasad komunikacji) oraz *afektywnej* (zawierającej poglądy na temat języka jako wartości kulturowej) (Maćkowiak 2020: 37–38). Należy zaznaczyć, iż kompetencja językowa użytkowników języka nie jest jednolita, a różnice w ocenie poprawności konstrukcji wynikają z takich czynników jak wiek, wykształcenie, płeć czy zawód, co nierzadko podważa skuteczność introspekcji jako jedynej metody badawczej. Eksperymenty pokazują, że bardziej skomplikowane i rzadziej używane konstrukcje są trudniejsze do jednoznacznej oceny, a intuicje językowe mogą być zniekształcone przez teoretyczne założenia badaczy (Гридина 2015: 172–179). W ten sposób przechodzimy do następnego kryterium, powiązanego z intuicją, a zatem do obiektywnego sprawdzania formułowanych opinii i ocen, czyli wykorzystania metod analitycznych.

Językoznawcy od dawna podkreślają konieczność stosowania metod empirycznych, takich jak analiza korpusowa czy metody statystyczne, by uwzględniać nie tylko zmienność społeczną czy demograficzną w badaniach nad językiem, ale obiektywniej dokonywać weryfikacji założonych hipotez (Ейгер, Безуглая, Бабич 2016: 165–166). Podstawową zaletą wykorzystania korpusów tekstowych jest przede wszystkim obiektywność zawartych w nich danych, które są oparte na rzeczywistych tekstach, co minimalizuje ryzyko subiektywizmu w interpretacji wyników. Badania korpusowe są nie tylko powtarzalne i weryfikowalne, co znacząco podnosi ich wiarygodność oraz naukową wartość, ale również umożliwiają prowadzenie analiz zarówno w kontekście pojedynczego języka, jak i porównawczych badań wielojęzycznych z uwagi na różnorodność zgromadzonych tekstów. Dostępne jednojęzyczne korpusy umożliwiają szczegółową analizę danego języka w różnych rejestrach, a także precyzyjne określenie semantyki wybranych jednostek (zarówno jedno-, jak i wielowyrazowych). Jednocześnie pozwalają na rozwiązywanie problemów związanych z wieloznacznością leksemów dzięki uwzględnieniu różnorodnych kontekstów, w których dana jednostka językowa występuje. Z kolei wielojęzyczne zasoby umożliwiają badanie różnic i podobieństw między językami, co jest szczególnie istotne w studiach kontrastywnych, przekładowczych czy leksykograficznych. W praktyce oznacza to możliwość badania takich zagadnień, jak frekwencja wybranych jednostek językowych, ich kolokacje, a także funkcjonowanie w różnych sytuacjach komunikacyjnych. Jednym z najważniejszych atutów

wykorzystania korpusów językowych jest możliwość precyzyjnego określenia częstości występowania określonych jednostek w naturalnych tekstach, co pozwala wyeliminować jednostki o charakterze okazjonalnym i skupić się na tych, które rzeczywiście są reprezentatywne dla danego języka lub rejestru.

Przykładem dziedziny, w której dokonano udanego połączenia podejścia intuicyjnego z metodami analizy korpusowej, jest lingwistyka czeska. Tradycyjnie oparta na intuicyjnych założeniach i ograniczonych badaniach sondowych, do połowy XX wieku charakteryzowała się przewagą subiektywnych metod badawczych. Przełom w tym zakresie przyniosło wprowadzenie analiz korpusowych, które umożliwiły empiryczne podejście do badania języka. Analizy te pozwoliły na głębsze zrozumienie uzusu językowego oraz na umiejscowienie intuicji badawczej w kontekście szerokich, obiektywnych danych. Jednym z pionierów tego podejścia był H. Kučera, który w swoich pracach podkreślał znaczenie analizy frekwencji oraz dynamiki uzusu językowego. Jego badania, prowadzone od lat pięćdziesiątych XX wieku, stały się podstawą dla późniejszych projektów empirycznych, zwłaszcza tych koncentrujących się na języku mówionym. Kučera, jako jeden z pierwszych, wprowadził systematyczne analizy statystyczne w badaniach lingwistycznych, co umożliwiło obiektywne porównania między intuicyjnymi założeniami lingwistów a rzeczywistym użyciem języka przez jego użytkowników (Sgall 2010: 86). Istotnym wkładem w rozwój badań korpusowych w Czechach były także projekty badawcze związane z Uniwersytetem Karola w Pradze, które skupiały się, m.in. na analizie języka mówionego w środowiskach miejskich, co zaowocowało stworzeniem frekwencyjnego słownika czeskiego języka mówionego, precyzyjnie ukazującego dynamikę języka codziennego. Czeskie badania korpusowe umożliwiły zatem przejście od subiektywnej intuicji do obiektywnego rozumienia języka. Współczesne podejście łączy dane korpusowe z intuicją rodzimych użytkowników, co nie tylko sprzyja realistycznej analizie i tworzeniu materiałów dydaktycznych odzwierciedlających naturalny uzus językowy, ale także stanowi cenne źródło doświadczeń i inspiracji dla lingwistów z innych krajów.

Lingwistyka jako w pewnym stopniu nauka intuicyjna, bazująca na wewnętrznym postrzeganiu i refleksji nad normami oraz regułami językowymi, wciąż nie wypracowała jednak w pełni rozwiniętej metodologii badawczej. Intuicja, choć kluczowa w opisie i analizie zjawisk językowych, niezaprzeczalnie wymaga wsparcia narzędziami umożliwiającymi obiektywizację danych i ograniczenie subiektywizmu badaczy (Čermák 1992: 19–20). Współczesne podejścia w lingwistyce coraz wyraźniej podkreślają konieczność równoważenia metod intuicyjnych, opartych na refleksji wewnętrznej, z empirycznymi podejściami opartymi na systematycznej analizie danych i obiektywnych obserwacjach, szczególnie przy wykorzystaniu korpusów

językowych. Integracja intuicji z obszernymi bazami korpusowymi oraz narzędziami analitycznymi umożliwia precyzyjną weryfikację formułowanych hipotez. Bazując na powyższych rozważaniach, podejmiemy próbę zaprezentowania różnych metod wyodrębniania wielowyrazowych jednostek fatycznych i emotywnych, w których kluczową rolę na pierwszym etapie analizy odgrywa intuicja badacza, natomiast podjęte decyzje zostaną zweryfikowane z wykorzystaniem dostępnych jedno- i wielojęzycznych korpusów tekstów w języku polskim, czeskim i rosyjskim. Na początku przedstawione zostaną metody dochodzenia do reproduktów od podstaw, czyli bez konieczności posiadania uprzednio zgromadzonej bazy wielowyrazowców. Dopiero w dalszej części omówione zostaną sposoby, które wymagają wykorzystania istniejącej bazy jednostek wielowyrazowych.

6.1 Metoda słownikowo-kolokacyjna

Podjęcie badań nad odtwarzalnymi jednostkami wielowyrazowymi musi zostać poprzedzone zebraniem reprezentatywnej bazy reproduktów, która w dalszych etapach analizy winna stanowić podstawę do dalszych poszukiwań interesujących nas wyrażeń bądź służyć jako materiał uzupełniający do zgromadzonych danych. Pierwszą z zaprezentowanych przez nas koncepcji dochodzenia do reproduktów wielowyrazowych można potraktować jako metodę wyjściową, pozwalającą ustanowić najczęściej używane jednostki języka mówionego (w naszym przypadku fatyzmy i emotywy) przy jednoczesnym „uruchomieniu” całego szeregu procedur naprowadzających, opartych w pierwszej kolejności na działaniu intuicyjnym badacza i jego wypracowanych kompetencji językowych. Sposób ten nazwijmy *metodą słownikowo-kolokacyjną*, polegającą na wykorzystaniu jednojęzycznego słownika skonstruowanego w porządku alfabetycznym do wnikliwej obserwacji zawartych w nim haseł, które mogą stanowić początek ciągu wielowyrazowego bądź jego element składowy. Innymi słowy, oznacza to śledzenie tych konstrukcji, w przypadku których osobiste poczucie językowe (odwołujące się do pamięci) sygnalizuje, że zaobserwowany wyraz prawdopodobnie występuje jako składnik większej całości o określonej odrębności kształtowo-znaczeniowej. W celu weryfikacji założonych hipotez należy zwrócić się do narodowych korpusów tekstów i zawartego w nich narzędzia kolokatora, umożliwiającego odnalezienie najbardziej łączliwych wyrazów dla konkretnej jednostki w określonym otoczeniu tekstowym. Po wyodrębnieniu ciągów wielowyrazowych z zawartych w korpusach poświadczeniach, proces końcowy stanowi określenie ich frekwencji i rozpowszechnienia w mowie użytkowników wskazanego języka w

oparciu o wybrane zasoby tekstowe. Opisaną metodę można przedstawić za pomocą poniższego schematu:



Za przydatne w danym procesie pytania, które umożliwiają dokonanie wstępnej selekcji ogromnej bazy jednostek słownikowych, mogą posłużyć sformułowania, typu: *W jakim wyrażeniu można zastosować dane słowo?*; *Jakie wyrazy występujące po danym słowie są najbardziej prawdopodobne?* czy też *W jakim kontekście dane słowo zostało wcześniej zaobserwowane?* Każdorazowe odwoływanie się do rejestru pamięci językowej zwiększa prawdopodobieństwo prawidłowego odnalezienia wyrażenia, w którym analizowana jednostka stanowi początek, część, koniec bądź wariant określonego wielowyrazowca. Przedstawmy egemplifikacyjnie przypadek obserwacji w jednojęzycznych słownikach języka polskiego, rosyjskiego i czeskiego artykułu hasłowego, odnoszącego się do wyrazu **koń/конь/kůň**, który – zdaniem autora (odwołującego się do pamięci frazematycznej) – stanowi element następujących reprodukowalnych jednostek wielowyrazowych:

KOŃ/KOŃ/KŮŇ
- No i zrobili nas w konia! - denerwowali się zgromadzeni na zebraniu mieszkańcy. - <i>To jakieś kłamstwa - oburzali się - Dwa lata temu były konsultacje w tej sprawie, ale wszyscy byli przeciwni.</i>
— <i>Вот и знают много, и вроде и понимают все на свете, а жить не умеют. А?</i> — <i>Да где они там знают много!</i> — <i>сказал я тоже со злостью. — Там насчет знаний-то... конь не валялся.</i>
<i>Největší radost mu udělal Milan Baroš , který svůj výkon korunoval krásným gólem. "To je můj kůň", rozdává radost. Škoda jenom, že nedostává příležitost od začátku utkání.</i>

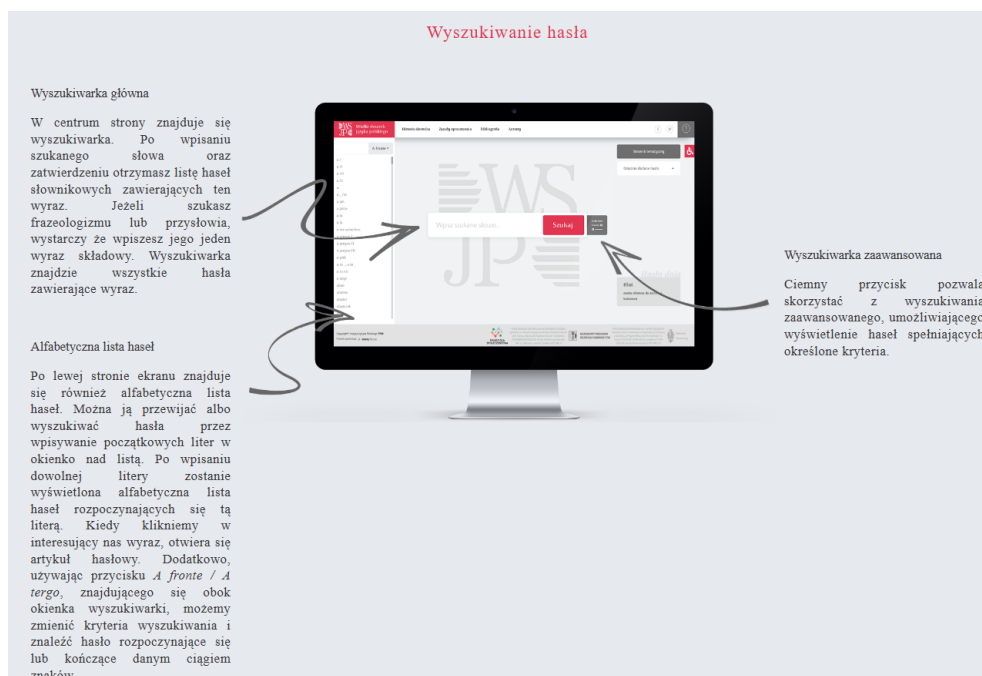
W danym przypadku pominięty zostaje proces zastosowania kolokatora do odnalezienia potencjalnych jednostek współwystępujących w bezpośredniej bliskości zadanego zapytania z powodu bezpośredniego przywołania (podczas obserwacji) wskazanych powyżej potocznych wielowyrazowców.

Pierwsze z wyodrębnionych wyrażen można ująć w słownikowej formie wyjściowej, tj. *ktoś robi w konia kogoś* z przypisaniem mu następującej definicji: *celowe wprowadzenie kogoś w błąd przy wykorzystaniu naiwności rozmówcy, przeważnie wynikające z potrzeby osiągnięcia własnych korzyści bądź żartowania z danej osoby*. Liczba poświadczeń dla danego wyrażenia w NKJP, uwzględniająca jego zastosowanie w różnych wariantach i kontekstach użycia, wyniosła 202 wyników. Rosyjskie wyrażenie *конь не валялся* stanowi przykład reproduktu leksykalnie zamkniętego, który sporadycznie bywa uzupełniony przez mówiącego elementami inicjalnymi, takimi jak <Здесь> *ещё и конь не валялся*. Owo wyrażenie służy podkreśleniu bierności działań w określonej sytuacji i odległej perspektywy ich realizacji, np.: —*Завтра свадьба, а у нас конь не валялся!* — *подтвердил Семён Иммануилович*. Korpus Araneum Russicum Maius, będący częścią zbioru Narodowego Korpusu Języka Czeskiego, rejestruje użycia danej frazy w 105 poświadczeniach. Z kolei odnotowany w niespełna 40 akapitach tekstowych czeskiego zbioru InterCorp v16ud zwrot *To je můj kůň* stosowany jest w ocenie kogoś lub czegoś, kto wzbudza zaufanie mówiącego i spełnia założone warunki do osiągnięcia konkretnego sukcesu. Z uwagi na niską liczbę zarejestrowanych poświadczeń, zasadnym byłoby odwołanie się do pozostałych zbiorów tekstowych w celu oceny stopnia rozpowszechnienia danego reproduktu w języku mówionym. W dalszej części analizy zaprezentowane zostanie praktyczne zastosowanie metody słownikowo-kolokacyjnej w oparciu o wybrane słowniki jednojęzyczne wraz z uwzględnieniem próbki uzyskanych wyników.

Reprodukty polskie: Elektroniczny słownik WSJP i kolokator NKJP

Wielki Słownik Języka Polskiego PAN (www.wsjp.pl) należy charakteryzować jako aktualnie największy słownik elektroniczny współczesnej polszczyzny, który przez twórców został określony jako słownik dokumentacyjny, czyli oparty na autentycznej bazie materiałowej, którą w pierwszej kolejności stanowią zbiory tekstowe NKJP (Żmigrodzki 2018: 17). Zaplanowana od samego początku forma elektroniczna słownika umożliwia obserwację

zawartych w nim haseł w porządku alfabetycznym bądź bezpośrednio wpisanie zapytania w wyszukiwarce głównej wraz z możliwością wybrania określonych kryteriów, które zostają przypisane dla wyszukiwanego hasła:



Konstrukcja zaprezentowanego słownika ma dostarczyć użytkownikowi możliwie najbardziej reprezentatywnej informacji dotyczącej wybranego artykułu hasłowego, która zawiera, m.in.: semantykę jednostki, zarejestrowane warianty i synonimy, przykłady użycia, kwalifikację tematyczną czy wzory odmiany. Szczegółowy opis opracowania słownika, jego możliwości i instrukcji użytkowania jest dostępny pod adresem <https://wsjp.pl/page/zasady> [data dostępu: 29.10.2024].

Cel naszego badania zawiera się w próbie odnalezienia haseł słownikowych, mogących stanowić część większej całości wielowyrazowej o określonej strukturze i znaczeniu przy wykorzystaniu:

- 1) obserwacji haseł, którą nazwiemy *poszukiwaniem intuicyjnym*;
- 2) alfabetycznego układu słownika WSJP;
- 3) narzędzia kolokatora NKJP;
- 4) zasobów tekstowych korpusów NKJP, KWJP (Korpus Współczesnego Języka Polskiego) i MoncoPL w celu weryfikacji znaczenia i frekwencji wyzyskanych jednostek.

Przebieg zastosowania przedstawionego modelu zostanie zilustrowany na przykładzie wyodrębnionego hasła **AFERA**.

afera

Hasło ma wiele znaczeń,
wybierz to, które Cię interesuje

1. obyczajowa

2. korupcyjna

CHRONOLOGIZACJA

FRAZEOLIZMY

POKAŻ WSZYSTKO

WSJP podaje następujące definicje dla wskazanej jednostki leksykalnej: 1) *Afera obyczajowa* – sensacyjne wydarzenie bulwersujące wielu ludzi; 2) *Afera korupcyjna* - ujawnione przedsięwzięcie o charakterze kryminalnym, w które zamieszane jest wiele osób. Ponadto użytkownikowi zostaje udostępniona informacja o możliwych występujących połączeniach, takich jak *wielka afera*; *afera miłosna*; *doszło do afery* itp. bądź

frazelogizmach, np. *ktoś gra na aferę*. Pomimo rozbudowanej struktury artykułu hasłowego nie otrzymujemy danych, dotyczących zastosowania wyrazu *afera* w reakcyjnych komunikatach potocznych. W tym celu odwołujemy się do narzędzia kolokatora NKJP i jego funkcji, który poprzez selekcję zbiorów tekstowych udostępnia najczęstsze połączenia wyrazowe, wśród których wybraliśmy następujące przykłady:

Kryteria ośrodka kolokacji:

afera**

Maksymalny odstęp: Zachowaj szyk: ☒

Kryteria kolokatu:

Część mowy: Kontekst z lewej: Kontekst z prawej:

Wielkość próbek: Min. współwystąpien:

Zaawansowane [URL](#)

SZUKAJ

[<< Poprzednie](#) [Następne >>](#)

7.	cała	cała <i>afera</i> (96), cały <i>afera</i> (1),	97	14,205.74
30.	być	jest <i>afera</i> (85), była <i>afera</i> (73), będzie <i>afera</i> (20), byłaby <i>afera</i> (14), być <i>afera</i> (8), było <i>afera</i> (3), był <i>afera</i> (2), są <i>afera</i> (1), będziemy <i>afera</i> (1),	207	943.15
37.	żaden	żadna <i>afera</i> (7), żadnym <i>afera</i> (2),	9	468.48
38.	dopiero	dopiero <i>afera</i> (17),	17	430.53
81.	raz	razu <i>afera</i> (4), raz <i>afera</i> (3), razem <i>afera</i> (2), razie <i>afera</i> (1),	10	25.81
34.	siebie	się <i>afera</i> (107), sobie <i>afera</i> (1),	108	197.81

Zastosowanie zapytania, uwzględniającego symbol dwóch gwiazdek, umożliwia wyszukiwanie fleksyjne słownikowe i tych konkordancji, które zawierają wszystkie formy odmian wskazanej jednostki. Następnie wyodrębnione przez korpus propozycje ośrodków kolokacyjnych zostały zweryfikowane w poszczególnych listach poświadczeń dla każdego z nich, wśród których udało się wyodrębnić następujące jednostki wielowyrazowe, podejrzewane o regularną odtwarzalność:

cała__afera	<i>stąd cała afera; I o to cała afera; No i to jest ta cała afera; Po co ta cała afera?</i>
jest__afera	<i>Jest afera; To jest dopiero afera</i>
żaden__afera	<i>To żadna afera</i>
dopiero__afera	<i>To dopiero byłaby afera!; To dopiero afera!</i>
raz__afera	<i>To od razu afera; I od razu afera</i>
siebie__afera	<i>Zrobiła się afera; Szykuje się afera</i>

Jak wynika z powyższej tabeli, niektóre z wyekscerpowanych reproduktów stanowiły wariant konkretnego wyrażenia, a różnica w ich strukturze dotyczyła wymienności bądź braku jednego czy więcej elementów, np. *To <jest> dopiero afera!*. Kolejnym etapem, poprzedzającym proces ustalenia stopnia frekwencji dla danych jednostek, jest próba określenia ich znaczenia semantycznego i pełnionej w wypowiedzi funkcji w oparciu o przytoczone przez korpus poświadczenia tekstowe:

<i>Stąd cała afera; I o to cała afera; No i to jest ta cała afera</i>
Komentarz podsumowujący. Używany w sytuacji: 1) Dla wskazania przyczyny powstania określonej sytuacji; 2) Dla skomentowania sytuacji, której powaga wydaje się być nieproporcjonalna w stosunku do rzeczywistego problemu.
1) Ładunek jest minimalny i nie zagraża życiu ani mieszkańców ani pracowników czy gości pałacu. Ale niewątpliwie zagraża interesom pałacu, <i>stąd cała afera.</i> 2) Ja myślę że szatan zakochał się w Ewie, bo była pierwszą kobietą <i>i o to cała afera.</i> A że Ewa miała Adama. to i jest co jest. 3) Więc zaczęły się te nieporozumienia, a więc ta pani zrobiła co zrobiła, czyli z dwóch metrów, po prostu stanęła, uruchomiła samochód, uderzyła w ścianę, wybuchły poduszki <i>no i to jest ta cała afera.</i>

<i>Po co ta cała afera?</i>
Komentarz inicjalny reakcyjny (na początku zdania) bądź podsumowujący (na końcu). Używany w sytuacji: Dla wyrażenia niezrozumienia wobec przesadnych (zdaniem mówiącego) emocji w stosunku do omawianego tematu.
1) Roman Rezmerowski, wójt gminy Tczew, mówi, że decyzję o warunkach budowy wydał, bo nie miał wyboru. Jest ona jednak nieprawomocna i mieszkańcy w ciągu 14 dni mogą się od

niej odwołać [...]. – **Po co ta cała afera?** – mówi jeden z mieszkańców Zabagna. 2) Jeżeli nie zrobią tego jednego dnia to zrobią kolejnego, więc **po co ta cała afera?**

Jest afera

Komentarz reakcyjny / oceniający jako składnik wypowiedzi. Używany w sytuacji: Dla określenia nagłego powstania nieprzyjemnego problemu, w wyniku którego uczestnicy rozmowy mogą ponieść pewne konsekwencje. Nierzadko służy do pobudzenia rozmówcy do podjęcia działań w celu zaniechania problemu.

1) – **Jest afera** - tłumaczył. - UOP szaleje, w Olsztynie i w Białymstoku od wczoraj wieczór siedzi dwóch podsekretarzy stanu. 2) I słyszę w słuchawce takie podenerwowane szept: - Słuchaj, dzwoni z policji jakiś facet, **jest afera**, gdzie jest ten Zdzichu?

To <jest> dopiero afera!

Komentarz podsumowujący. Używany w sytuacji: Dla wyrażenia zdumienia wobec omawianej sytuacji wraz z podkreśleniem powagi i kontrowersji podejmowanego problemu. Stosowany przeważnie w ironicznym bądź prześmiewczym tonie.

1) W Pruszkowie powstał tor kolarski za 90 milionów złotych. I co? I nic. Dopiero teraz w związku kolarskim zaczynają myśleć, jak ten tor wykorzystać, jak zapelnić go zawodnikami. A w Polsce na torze jeździ 100 cyklistów. **To jest dopiero afera!** 2) Nie minęło wiele czasu, a PSL opchnął tę nieruchomość spółce deweloperskiej za 72 miliony [...]. Proszę, **to jest dopiero afera** [...]

To żadna afera

Komentarz reakcyjny (samodzielny)/ oceniający jako składnik wypowiedzi / podsumowujący. Używany w sytuacji: Dla wyrażenia zdziwienia wobec sytuacji, która zdaniem mówiącego jest niewarta zamieszania. Fraza stosowana również w celu zminimalizowania powagi problemu.

1) **To żadna afera** – mówił wczoraj wicedyrektor placówki dr Mariusz Witczak. 2) Oglądając film pomyślałem, że **to żadna afera**, bo skoro pracuje tam ciężki sprzęt to pewnie wygrodzony [...] 3) Fakt, że Polacy mają do tego krytyczny stosunek, ale trzeba mieć świadomość, że nie jest **to żadna afera**.

<i>To dopiero byłaby afera</i>
Komentarz inicjalny reakcyjny (na początku zdania) bądź podsumowujący (na końcu): Używany w sytuacji: W odniesieniu do wyobrażanego problemu i jego okoliczności, którego konsekwencje mogłyby być poważne lub kontrowersyjne, jeśli by się wydarzyły. Może pełnić funkcję przestrogi.
1) <i>To dopiero byłaby afera, gdybyśmy najpierw położyli nowy asfalt, a po kilku miesiącach musielibyśmy go zrywać</i> 2) <i>W pobliżu dyskutują pracownicy. – Jakby ta kontrola przyjechała dzisiaj, to dopiero byłaby afera.</i>

<i>To od razu afera</i>
Komentarz podsumowujący. Używany w sytuacji: Dla wyrażenia niezrozumienia z powodu reakcji wobec problemu, który zdaniem mówiącego nie jest na tyle poważny.
1) <i>W Niemczech nikogo nie dziwi widok kanclerz pijącej na spotkaniu piwo, a jak ktoś u nas wypije w restauracji, to od razu afera.</i>

<i>I od razu afera</i>
Komentarz podsumowujący. Używany w sytuacji: Dla wyrażenia zaskoczenia i irytacji dla powtarzających się przesadnych reakcji w stosunku do omawianego problemu.
1) <i>A potem powiesz złe słowo na takiego Piłsudskiego którego osobiście delikatnie mówiąc szacunkiem nie darzę a wręcz przeciwnie jest to jeden z Polaków za których się wstydzę i od razu afera.</i>

<i>Zrobiła się afera</i>
Komentarz reakcyjny (samodzielny)/ oceniający jako składnik wypowiedzi / podsumowujący. Używany w sytuacji: Dla wskazania, że jakieś niespodziewane wydarzenie, które początkowo wydawało się być niewielkim problemem, wzbudziło silne emocje.
1) <i>Zrobiła się afera: pisowski bojówkarz uważa że homoseksualiści to zwierzęta i wysłannicy diabła.</i> 2) - <i>Zabrali mnie do szpitala, zrobiła się afera , bo jeszcze Maciuś wypelzł z kieszeni, kiedy znalazłem się na sali.</i> 3) <i>Policja traktowała to jako zniszczenie mienia niewielkiej wartości. Dopiero teraz, gdy człowiek ten rzucił się na policjanta, zrobiła się afera - wytyka nadleśniczy.</i>

<i>Szykuje się afera</i>
Komentarz inicjalny. Używany w sytuacji: Dla podkreślenia zbliżającego się problemu i jego pierwszych oznak, który najprawdopodobniej spotka się z silną i negatywną reakcją.
1) Szykuje się afera. <i>Losowanie drużyn przed Mistrzostwami Europy 2012 r. miało się początkowo odbyć w Teatrze Wielkim, ale prezes [...] zdecydował, że odpowiedniejszym miejscem będzie Sala Kongresowa w Pałacu Kultury.</i>

Jak można zauważyć, wyekscerpowane przykłady reproduktów wielowyrazowych z elementem „afera” stanowiły w przeważającej większości komentarz podsumowujący, sytuujący się na końcu wypowiedzi. Udostępnione przez narzędzie otoczenia tekstowe, zawierające interesujący nas ciąg wielowyrazowy, umożliwiają ustanowienie wszystkich znaczeń reproduktu w zależności od zaprezentowanej sytuacji komunikacyjnej. Wśród zbiorów tekstowych wybrano i opisano semantycznie wyłącznie te połączenia leksykalne, które w naszej ocenie można zakwalifikować do grupy emotywnów reakcyjnych, oceniających bądź podsumowujących. W związku z powyższym analiza nie uwzględniała przypadków użycia wskazanych wielowyrazowców w znaczeniu dosłownym (nie dających się charakteryzować jako wyrażenia ematywne), np.

<i>W swobodnej wypowiedzi stwierdziła, że działalność komisji jest niezbędna do zweryfikowania, czym właściwie jest afera <u>wizowa</u>, a czym nie jest.</i>
<i>Najlepszym tego przykładem jest „afera <u>Rywińna</u>” i komisja śledcza, która jest jedną wielką szopką.</i>
<i>Choć organizacje antykorupcyjne od lat ostrzegały przed niejasnymi zasadami pracy lobbystów w instytucjach UE, to dopiero afera <u>po publikacji</u> „Sunday Times” stworzyła [...] polityczną atmosferę.</i>

Końcowa część analizy zostaje poświęcona ustaleniu frekwencji dla każdego z wielowyrazowców, w przypadku których zaprezentowana poniżej liczba poświadczeń stanowi łączną sumę wszystkich konkordancji, uwzględnionych w korpusach NKJP, KWJP i Monco:

REPRODUKT	FREKWENCJA
<i>Stąd cała afera</i>	21 poświadczeń
<i>I o to cała afera</i>	4 poświadczenia
<i>No i to jest ta cała afera</i>	1 poświadczenie
<i>Po co ta cała afera</i>	15 poświadczeń

<i>Jest afera</i>	2164 poświadczenia
<i>To jest dopiero afera!</i>	23 poświadczenia
<i>To żadna afera</i>	11 poświadczeń
<i>To dopiero byłaby afera!</i>	15 poświadczeń
<i>To dopiero afera!</i>	18 poświadczeń
<i>To od razu afera</i>	7 poświadczeń
<i>I od razu afera</i>	16 poświadczeń
<i>Zrobiła się afera</i>	344 poświadczenia
<i>Szykuje się afera</i>	57 poświadczeń

Powyższa tabela i jej wyniki uwidaczniają silną dysproporcję, dotyczącą stopnia rozpowszechnienia wskazanych reproductów i ich zarejestrowania w poświadczeniach tekstowych, co nie powinno być powodem przedwczesnego odrzucenia jednostek mających niższą frekwencję, a raczej – zwrócenia się ku większym zasobom tekstowym w celu ich dalszej weryfikacji. Łącznie dla wszystkich reproductów wyodrębniono dokładnie 2696 konkordancji, wśród których przybliżony odsetek przykładów tekstowych poszczególnych korpusów przedstawia się następująco:

1. MoncoPL – 98% (2696 poświadczeń);
2. NKJP – 2% (50 poświadczeń);
3. KWJP – 1% (5 poświadczeń).

Uzyskane wyniki skłaniają do wysunięcia wniosku, iż korpus MoncoPL (codziennie rejestrujący teksty internetowe czy wypowiedzi internautów) okazał się najbardziej przydatną i reprezentatywną bazą tekstową dla wyodrębnienia i analizy jednostek języka mówionego, utrwalonych w zapisie elektronicznym. Ponadto należy podkreślić, że w przypadku jednostek o najwyższej frekwencji użycia (np. *jest afera* – 2164 poświadczenia) udostępnione przez korpus konkordancje uwzględniały wszystkie wcześniej zidentyfikowane znaczenia danej jednostki leksykalnej, przy czym liczba wystąpień dla każdego z nich sięgała dziesiątek przykładów. Poniżej zaprezentowano spis innych reproductów wielowyrazowych (fatyzmów i emotywów), które w sposób analogiczny zostały wyodrębnione i semantycznie zweryfikowane przy zastosowaniu metody słownikowo-kolokacyjnej:

WYRAZ	POŁĄCZENIA	REPRODUKTY WIELOWYRAZOWE
<i>Bóg</i>	<i>Bóg__raczy; Bóg__zapłać; Bóg__świadkiem</i>	<i>Bóg jeden raczy wiedzieć; Bóg raczy wiedzieć; Bóg zapłać; Bóg mi świadkiem</i>
<i>Chwila</i>	<i>chwilę__czasu; chwila__chwila</i>	<i>Daj mi chwilę czasu; Chwila, chwila!</i>
<i>Głowa</i>	<i>zawracaj__głowy; spokojna__głowa</i>	<i>Nie zawracaj mi głowy; Spokojna Twoja głowa</i>
<i>Jaja</i>	<i>robić__jaja; niezłe__jaja; dla__jaj</i>	<i>Pan sobie robi jaja?; Ty chyba sobie robisz jaja; No to niezłe jaja!; Na poważnie czy dla jaj?</i>
<i>Koniec</i>	<i>jeszcze__koniec; koniec__końców</i>	<i>To jeszcze nie koniec!; Koniec końców...</i>
<i>Licho</i>	<i>pal__lichu; lichu__porwie; u__licha</i>	<i>Pal lichu!; Niech Cię lichu porwie!; Co u licha?</i>
<i>Margines</i>	<i>na__marginesie; marginesie__dodam</i>	<i>Tak na marginesie...; Na marginesie dodam...</i>
<i>Prawda</i>	<i>Prawdę__mówiąc</i>	<i>Prawdę mówiąc...; Mówiąc prawdę...</i>
<i>Raz</i>	<i>każdym__razie; sam__raz</i>	<i>W każdym razie...; w sam raz</i>
<i>Słowo</i>	<i>Słowo__honoru; słów__brak,</i>	<i>Słowo honoru; brak mi słów</i>

Reprodukty rosyjskie: *Большой толковый словарь современного русского языка* i kolokator Narodowego Korpusu Języka Rosyjskiego

Próba odnalezienia odtwarzalnych wielowyrzawowych jednostek języka rosyjskiego została przeprowadzona w oparciu o *Большой толковый словарь русского языка* pod redakcją S. A. Kuzniecowa. z 2000 roku. Wykorzystane w analizie opracowanie leksykograficzne zawiera około 130 000 słów, dla których szczegółowo skonstruowany artykuł hasłowy dostarcza informacji, dotyczących pisowni wyrazu, miejsca akcentu, możliwych słów pochodnych czy sfery funkcjonowania wyrazu. Głównym kryterium doboru słownictwa było jego rzeczywiste i frekwencyjne użycie w tekstach beletrystycznych, publikacjach popularnonaukowych, w publicystyce, periodykach masowych czy języku mówionym. Należy również wspomnieć o bogatym repertuarze kwalifikatorów, którymi opatrzone zostaje każde hasło w słowniku wraz z przytoczeniem przykładu użycia leksemu/połączenia w zdaniach czy fakt włączenia do pracy zbioru skrzydlatych wyrażen (przysłów, powiedzeń, aforyzmów i cytatów uznanych klasyków literatury rosyjskiej) (Кузнецов 2000: 3). Znaczna przydatność danego opracowania leksykograficznego w lingwistyce stosowanej i badaniach nad odtwarzalnymi jednostkami wielowyrzawowymi znajduje potwierdzenie w interpretacji przez

autora przenośnych znaczeń wyrazów wraz z przytoczeniem stabilnych połączeń wielowyrazowych i jednostek frazeologicznych, np.

ВСЯЧИНА, -ы; ж. Разг. Разнородная смесь; всё, что угодно. На рынке продают разную всячину. ◇ Всякая всячина. 1. Смесь разнородных предметов, продуктов. Комната завалена всякой всячиной. 2. О чём-л. разнообразном, различном. Слушать всякую всячину. Со всячиной, в зн. нареч. По-разному, так и сяк. За последние годы жить пришлось со всячиной.

Od zaprezentowania możliwości niniejszego opracowania leksykograficznego postaramy się w sposób analogiczny ustanowić odtwarzalne wielowyrazowe jednostki języka rosyjskiego, wykorzystując zawarte w słowniku hasła. Nie ulega wątpliwości fakt, że efektywność poszukiwania intuicyjnego podczas analizy języka nierodzimego jest istotnie ograniczona. Jej skuteczność zależy w głównej mierze od stopnia znajomości połączeń frazematycznych, charakterystycznych dla danego języka, a w przypadku ich braku wymaga natomiast więcej czasu na określenie możliwych kolokacji dla wybranych jednostek.

Przedstawione badanie zostanie przeprowadzone na podstawie zaprezentowanego schematu:

- 1) obserwacja haseł w porządku alfabetycznym w słowniku *Большой толковый словарь русского языка*;
- 2) zastosowanie narzędzia kolokatora NKJR;
- 3) wykorzystanie zasobów tekstowych korpusów NKJR, rosyjskich zbiorów InterCorpv16ud i korpusu Araneum Russicum Maius w celu weryfikacji semantyki i frekwencji wyodrębnionych jednostek.

W naszym przykładzie posłużymy się hasłem 'ШУТКА', którego artykuł hasłowy w skróconej formie przedstawia się następująco:

ШУТКА, -и; мн. род. -ток, дат. -ткам; ж. 1. То, что говорят, делают с целью вызвать смех; забавная ШУТКА весёлая проделка, выходка или острота. Остроумная, тонкая ш. Невинная, безобидная ш. Злая, колкая ш. Обидеть кого-л. глупой шуткой. *Он шуток не понимает. Мне сегодня не до шуток.* Отпускать шутки направо и налево. 2. То, что говорят, делают не всерьёз, ради игры, забавы, развлечения. *Её плач был шуткой.* 3. Небольшая комическая пьеса. *Сыграть шутку (см. Сыграть). Шутки (шутку) шутить* (Кузнецов 2000: 1509).

Jak można zauważyć, część zaprezentowanego hasła dostarcza potencjalnie możliwych połączeń wyrazowych, wśród których autor wymienia kolejne: *шутки в сторону; шутки плохи с кем-чем; шутка шуткой; в шутку; не шутка; не до шуток; не в шутку; шутка ли* itp. Pomimo obecności gotowych konstrukcji językowych zawartych w słowniku, zwracamy się do narzędzia kolokatora NKJR i zmienności parametru *Расстояние* w celu wyodrębnienia kolokacji, które przedstawiono poniżej:

Ключ

Лемма 🔍 ?

добавить свойство ▲

Коллокат

Лемма 🔍 ? ✕

Грамм. признаки выбрать ? ✕

Расстояние
 от: до:

добавить свойство ▲

Ключ	Коллокат	Совместная частота	Частота ключа	Частота коллоката	LogDice ▼	Loglikelihood ▼	MI ³ ▼	t-score ▼	Агр. мера ▼	Конкорданс
шутка	без	440	24015	399192	7.82	1681.65	15.02	19.76	17.22	Примеры
шутка	шутка	560	24015	24015	10.24	5512.81	18.55	23.60	20.92	Примеры
шутка	не	5650	24015	6545609	7.63	20175.67	19.88	69.58	37.19	Примеры
шутка	это	1964	24015	2097072	7.71	7018.28	17.85	41.28	27.14	Примеры
шутка	в	4366	24015	11335734	6.83	8977.52	18.56	55.07	31.97	Примеры
шутка	ли	963	24015	454761	8.48	4906.41	17.24	30.09	22.78	Примеры

Kolejny etap analizy stanowi sprawdzenie wyodrębnionych przez korpus propozycji ośrodków kolokacyjnych, dla których konieczna pozostaje ich szczegółowa weryfikacja w oparciu o indywidualne listy poświadczeń przypisanych do każdej z nich. W trakcie weryfikacji zidentyfikowano następujące jednostki wielowyrazowe, które przejawiają potencjalną tendencję do regularnej odtwarzalności w badanych kontekstach:

шутка_без	<i>Без шуток; Говоря без шуток...; Я без шуток; если без шуток</i>
шутка_шутка	<i>Шутка шуткой...; шутки шутками...</i>
шутка_не	<i>Тут не до шуток</i>
шутка_это	<i>Это, конечно, шутка; Это такая шутка?</i>

шутка_в	<i>Шутки в сторону; шутка в том...; в каждой шутке есть доля правды.</i>
шутка_ли	<i>Шутка ли</i>

Niektóre z uzyskanych wyników pokrywają się z zaproponowanymi w słowniku połączeniami wyrazowymi, takimi jak *шутки в сторону; шутка шуткой; шутка ли* czy *<Тут> не до шуток*. Można również wstępnie założyć, że obecne w tabeli podobieństwa form strukturalnych wielowyznaczników *Шутка шуткой/шутки шутками* stanowią warianty produktu, które niewątpliwie charakteryzują się odmiennym stopniem frekwencji we wspomnianych zasobach korpusowych. Przystąpmy zatem do analizy, której celem jest określenie potencjalnych znaczeń semantycznych wyodrębnionych połączeń oraz zbadanie funkcji, jakie pełnią w przytoczonych poświadczeniach tekstowych:

<i>Без шуток</i>
Кomentarz reakcyjny (na początku zdania). Używany w sytuacji: 1) Dla stanowczego przerwania żartobliwego tonu wypowiedzi rozmówcy wraz z wyrażeniem irytacji ze strony mówiącego 2) W celu łagodnego wyrażenia prośby w stosunku do rozmówcy by wypowiedział się w kwestii określonego zagadnienia szczerze i poważnie.
<i>1) - Особенно опасайтесь девочку. У нее за ушами пулеметы. -А ну, без шуток! - обиделся адмирал, который не выносил, если его подозревали в трусости 2) Ничего я не задумал. Это я так, позлить вас. Такой же советский человек, как и вы. Меня недоверие обижает. — Так давайте, Греков, без шуток.</i>

<i>Говоря без шуток</i>
Кomentarz wyjaśniająco-podsumowujący. Używany w sytuacji: W celu przerwania żartobliwego tonu wypowiedzi i wyrażenia głównej myśli mówiącego w sposób poważny i pozbawiony ironii.
<i>1) Не так давно в поле моего зрения попала Фаня Долицкая. Попала и осталась там, никак не хочет исчезать...А говоря без шуток, эта миниатюрная арабистка мне очень нравится.</i>

<i>Я без шуток</i>
Кomentarz reakcyjny (na początku zdania). Używany w sytuacji: Dla podkreślenia powagi słów mówiącego (w stanowczym tonie), które początkowo przez rozmówcę mogą zostać potraktowane lekceważąco.
1) — <i>На основании чего трудимся? — На основании желания! — встрял Одутловатов. — Я без шуток!</i> — жестко сказал Харченко. — <i>Где документация?</i>

<i>Если без шуток</i>
Кomentarz wyjaśniająco-podsumowujący. Używany w sytuacji: Jako wtrącenie, wyprowadzające wypowiedź z jej żartobliwego charakteru i przejście do głównego wątku zagadnienia, który należy traktować poważnie.
1) <i>Агрессивность — знак того, что ты не в силах жить в безграничном, меняющемся, прекрасном и жестоком мире, космическом, если без шуток, а просто хочешь огородить вокруг себя уютненькое пространство [...].</i>

<i>Шутка шуткой/ Шутки шутками</i>
Кomentarz reakcyjny. Używany w sytuacji: Dla zakończenia humorystycznego wątku rozmowy w celu skierowania uwagi rozmówcy na poważniejszy aspekt omawianego zagadnienia.
1) <i>Шутка шуткой, а борьба за приоритет здесь показательна — раз друг к другу придираются, значит, повод действительно существует. 2)</i> <i>Шутки шутками, но я больше не сомневался ни в жизненной цели, ни в её достижимости.</i>

<i>Тут не до шуток</i>
Кomentarz reakcyjny (samodzielny)/emfaticzny (w środku wypowiedzi). Używany w sytuacji: 1) Dla przerwania ironicznego tonu wypowiedzi rozmówcy 2) Dla wyrażenia osobistego stosunku mówiącego w odniesieniu do podejmowanego tematu i podkreślenia powagi sytuacji
1) — <i>Ага, понимаю, ты свои деньжата спустил в казино и сейчас явился сюда за добавочкой, — с иронией отозвался майор. — Тут не до шуток,</i> — оборвал Афоня Карельский 2) <i>Тем не менее, у некоторых лиц проблема карьеры принимает драматический характер, тут не до шуток, если уже в детском возрасте наблюдается множественное разрушение зубов.</i>

<i>Это, конечно, шутка</i>
Кomentarz wyjaśniający (w środku wypowiedzi). Używany w sytuacji: Dla podkreślenia żartobliwego charakteru wcześniejszej wypowiedzi mówiącego, która nie wyraża jego rzeczywistych poglądów, i zminimalizowania ryzyka odebrania słów w sposób dosłowny.
1) <i>Например, он мог бы таскать за волосы судей или других тренеров. Но это, конечно, шутка. Если говорить серьезно, можно назначать кого угодно, ситуация в нашем футболе не изменится[...]</i>

<i>Это такая шутка?</i>
Pytanie reakcyjne. Używane w sytuacji: Dla wyrażenia niezrozumienia, nieufności bądź ironii w stosunku do otrzymanej od rozmówcy informacji. Może być użyte jako pytanie retoryczne (wyrażające negatywny pogląd mówiącego) bądź jako sygnał do bardziej szczegółowego wyjaśnienia wypowiedzianych słów przez rozmówcę.
1) <i>Вы должны держать ее в отеле, чтобы она была в безопасности. - В безопасности? Это такая шутка?</i>

<i>Шутки в сторону</i>
Кomentarz reakcyjny. Używany w sytuacji: W celu przzerwania żartobliwego tonu rozmówcy i rozpoczęcia poważniejszej dyskusji. Nierzadko służy podkreśleniu irytacji mówiącego z powodu lekceważącego stosunku rozmówcy do określonego zagadnienia.
1) <i>Тебе нужна атомная бомба? Мне — не нужна. Я, как и Земеля, к мировому господству не стремлюсь. — Но шутки в сторону! — спохватился опять Рубин</i>

<i>Шутка в том</i>
Кomentarz wyjaśniający (w środku wypowiedzi). Używany w sytuacji: Jako wprowadzenie do wyjaśnienia absurdalnego bądź ważnego szczegółu sprawy, który nie jest znany bez informacji, przekazanej w dalszej części wypowiedzi mówiącego.
1) <i>Клоуны делали репризу, которая традиционно заканчивалась тем, что на зрителя падает какой-то шест, на котором ведро с водой... Шутка в том, что воды никакой нет, а зритель — подсадка, разумеется.</i>

<i>В каждой шутке есть доля правды</i>
Кomentarz testimonialny (jako wtrącenie). Używany w sytuacji: Dla podkreślenia ukrytej prawdziwości wyrażonego stwierdzenia wraz z możliwym odzwierciedleniem rzeczywistej opinii mówiącego w określonym temacie.
1) <i>Теперь же художник говорит в шутку (хотя в каждой шутке есть доля правды), что все диваны знаменитого француза ужасно неудобны, и он отправил их в мастерскую!</i>

<i>Шутка ли</i>
Кomentarz emfaticzny (jako wtrącenie). Używany w sytuacji: 1) Dla podkreślenia prawdziwości omawianego zagadnienia w sposób łagodny 2) Dla wyrażenia powagi prezentowanego zagadnienia i wywołania emocjonalnej reakcji u odbiorcy.
1) <i>Помню, как мы с женой шифоньер на трамвае везли, шутка ли, от самого магазина, а потом на новой «копейке» чугунную ванну. 2) Основные проблемы[...] - это низкий платежеспособный спрос населения и, соответственно, низкая активность населения на потребительском рынке [...] -Шутка ли, если реальная доля населения, активно участвующего на потребительском рынке, сегодня колеблется, по разным оценкам, от 30 до 40 %!</i>

Zdecydowana większość zaprezentowanych wielowyrazowców, uwzględniających leksem *шутка*, stanowiła w danym przypadku wtrącenie w środku wypowiedzi, które winno pełnić funkcję wyjaśniającą w stosunku do znajdującej się przed reproduktorem informacji bądź emfaticzną, mającą na celu intensyfikację siły przekazu ze strony nadawcy. Podjęta analiza, podobnie jak w przypadku prób wyodrębnienia jednostek polskich, ponownie wykazała trudności w ekscerpacji interesujących nas wyrażen emotywnych z powodu obszernych zastosowań niektórych jednostek wyłącznie w znaczeniach dosłownych, np. *без шуток*:

<i>Он излагал мысли быстро, внятно, энергично: обходился без шуток и каскадов метафор, апеллировал к логике, игнорировал тех, кто его перекрикивал.</i>
<i>Он зашел к ней спокойно, без шуток и прибауток: понял уже, что Татьяну веселые заигрывания отпугивают.</i>
<i>Говорил без шуток, совершенно серьезно, так что все его прихожанки ходили чинно, в юбках до пят.</i>

Problemy w procesie ustalania semantyki napotkano również w przypadku połączenia wielowyrazowego *Это такая шутка?*, którego wpisanie w korpus przynosiło także konkordancje, zawierające tę konstrukcję w trybie oznajmującym, które naturalnie posiadało już inne znaczenie w wypowiedzi (NKJR pominął wskazany znak interpunkcyjny). W końcowej części badania oszacowano łączną liczbę poświadczeń tekstowych dla każdego wielowyrazowca przy użyciu zasobów korpusów NKJR, rosyjskiego podkorpusu InterCorp v16ud i Araneum Russicum Maius:

РЕПРОДУКТ	FREKWENCJA
<i>Без шуток</i>	88 poświadczeń
<i>Говоря без шуток</i>	5 poświadczeń
<i>Я без шуток</i>	27 poświadczeń
<i>Если без шуток</i>	108 poświadczeń
<i>Шутка шуткой</i>	30 poświadczeń
<i>Шутки шутками</i>	413 poświadczeń
<i>Тут не до шуток</i>	35 poświadczeń
<i>Это, конечно, шутка</i>	66 poświadczeń
<i>Это такая шутка?</i>	25 poświadczeń
<i>Шутки в сторону</i>	396 poświadczeń
<i>Шутка в том</i>	28 poświadczeń
<i>В каждой шутке есть доля правды</i>	76 poświadczeń
<i>Шутка ли</i>	1 246 poświadczeń

Представленные результаты badań w sposób jednoznaczny raz jeszcze dowodzą zróżnicowanego poziomu występowania oraz upowszechnienia wyodrębnionych reproduktów w analizowanych tekstach, które zostały uwzględnione w badanych korpusach. Otrzymane dane wskazują na istotne różnice w częstości oraz kontekście użycia analizowanych reproduktów, co pozwala na wyciągnięcie wniosków o ich specyficznej funkcji w poszczególnych typach tekstów. Łączna liczba wyekscerpowanych konkordancji dla wszystkich reproduktów wyniosła 2543 poświadczenia, a przybliżony udział przykładów tekstowych z poszczególnych korpusów kształtuje się następująco: 1. NKJR – 44% (1119 poświadczeń); 2. Rosyjski moduł InterCorp v16ud – 7% (174 poświadczenia); 3. Araneum Russicum Maius – 49% (1250 poświadczeń). Jak wynika z danych liczbowych, NKJR i porównywalny korpus Araneum Russicum Maius wykazują zbliżony stopień ich użyteczności

wykorzystania w procesie poszukiwania i identyfikacji jednostek wielowyrazowych potocznych. Wynika to przede wszystkim z wysokiej liczby ich zasobów tekstowych, tj. NKJR – 2 miliardy słów, A.R.M – 1,2 miliarda słów, przy czym rosyjskie zasoby korpusu InterCorp v16ud rejestrują zaledwie ok. ponad 110 mln. Poniżej przedstawiono spis przykładowych reproduktów wielowyrazowych, zawierający zarówno fatyzmy, jak i emotywy, które zostały analogicznie wyodrębnione oraz poddane semantycznej weryfikacji przy użyciu metody słownikowo-kolokacyjnej:

WYRAZ	POŁĄCZENIA	REPRODUKTY WIELOWYRAZOWE
<i>абзац</i>	<i>абзац_полный</i>	<i>Это полный абзац!</i>
<i>говорить</i>	<i>говорить_о; говорить_не; говорить_что</i>	<i>О чём тут говорить; не тебе говорить; Что тут говорить..</i>
<i>дело</i>	<i>дело_самый; дело_это; дело_не</i>	<i>На самом деле; другое дело, что...; дело не в этом</i>
<i>значение</i>	<i>значение_иметь; значение_придавать</i>	<i>Это не имеет значения; Не придавайте значения</i>
<i>мало</i>	<i>мало_тот; мало_сказать</i>	<i>Мало того...; мало сказать...</i>
<i>пора</i>	<i>пора_давно; пора_валить; пора_ли</i>	<i>Тебе давно пора; пора валить отсюда; Не пора ли...</i>
<i>сторона</i>	<i>сторона_другой; сторона_тот; сторона_один</i>	<i>С другой стороны...; По ту сторону взгляда...; с одной стороны...</i>
<i>точка</i>	<i>точка_с; точка_в; точка_на</i>	<i>С точки зрения...; в точку! Точка на этом</i>
<i>чёрт</i>	<i>чёрт_побрать; чёрт_знать; чёрт_какой</i>	<i>Чёрт побери! Чёрт знает! Какого чёрта?</i>
<i>язык</i>	<i>язык_зуб; язык_снять</i>	<i>Держи язык за зубами; с языка сняли</i>

Reprodukty czeskie: *Slovník současné češtiny* LINGEA i kolokator Narodowego Korpusu Języka Czeskiego

Odnalezienie i identyfikacja odtwarzalnych, wielowyrazowych jednostek języka czeskiego została przeprowadzona przy użyciu pracy *Slovník současné češtiny*, będącego opracowaniem zbiorowym autorstwa zespołu badaczy współpracujących z wydawnictwem Lingea, który opublikowano w 2011 roku. To opracowanie leksykograficzne zawiera zbiór około 30 000 haseł, uznanych za reprezentatywne i charakteryzujących się najwyższą frekwencją występowania w języku zarówno w kontekstach sfery publicznej, jak i prywatnej.

Wyselekcjonowane jednostki leksykalne odzwierciedlają kluczowe tendencje językowe, umożliwiając wnikliwą analizę struktury i użycia jednostek w różnych rejestrach i sytuacjach komunikacyjnych. Zgodnie z zamysłem autorów, słownik ma pełnić funkcję wszechstronnego narzędzia wspomagającego pracę z tekstem, ułatwiającego precyzyjne wyrażanie intencji i myśli w codziennej komunikacji. Każdy artykuł hasłowy dostarcza kompleksowej informacji na temat znaczenia danego wyrazu, jego poprawnej pisowni, typowych kolokacji z innymi leksemami czy charakterystyki stylistycznej. Istotnym i wartym podkreślenia zabiegiem zastosowanym jest uwzględnienie nie tylko standardowego języka literackiego, lecz także współczesnych odmian języka, w tym czeszczyznę potoczną (*obecná čeština*) oraz rozmaite slangi. O nowatorskim charakterze słownika oraz jego wysokiej użyteczności dla szerokiego grona odbiorców, w tym również osób spoza środowiska akademickiego, świadczą zastosowane rozwiązania graficzne, takie jak kolorystyczne wyróżnienie poszczególnych kategorii informacji dotyczących danej jednostki leksykalnej oraz dodatkowe ilustracje dla wybranych haseł. Szczególnie istotnym elementem badań nad reproduktami jest uwzględnienie w niektórych hasłach słownikowych idiomów (autorzy słownika określają je frazemami), co umożliwia ekscerpcję ustalonych konstrukcji językowych i ich dalszej weryfikacji w korpusie wraz z poszukiwaniem ekwiwalentów polskich:

ani ¹ spoj. souř. slučuje v záporné větě dva nebo více větných členů stejné platnosti nebo dvě i více souřadných záporných vět (před ani se píše čárka zprava, jen při zdůraznění druhého členu a ve spojení ani – ani) ▷ Dva dny nejedl ani nepil. • V zimě ani nebruslí, ani nelyžuje. • Nemá ani sestru, ani bratra.

ani ² část. □ zdůrazňuje, že se zápor týká také případu, o kterém by se to nepředpokládalo ▷ Nevěřím mu ani slovo. • Ať tě to ani nenapadne! • Ani se neptej! • ani za mák = vůbec • ani nápad = to nepřipadá v úvahu • ani co by se za něhet vešlo = vůbec nic □ vyjadřuje neurčitost, nerozhodnost ▷ Půjdeš se mnou? – Ani nevím. • Chceš napít? Ani ne.

chápat ned. (1. j. -u; 3. m. -ou; rozk. -ej) □ postihovat rozumem □ rozumět ▷ chápat pomalu/rychle/snadno/lehce • chápat smysl slov □ mít pochopení, porozumění ▷ Nechápu, co po mně chcete. • Chápu vaše rozhodnutí.

W ramach niniejszego opracowania leksykograficznego podejmiemy próbę wyodrębnienia i klasyfikacji odtwarzalnych wielowyrazowych jednostek języka czeskiego, bazując na analogicznej do poprzednich analiz procedurze:

- 1) obserwacja haseł w porządku alfabetycznym w pracy *Slovník současné češtiny*;
- 2) zastosowanie funkcji *Kolokace* w korpusie SYN v8;
- 3) wykorzystanie zasobów tekstowych korpusów NKJCZ (Syn v8, ORAL v1 i Araneum Bohemicum Maximum) w celu weryfikacji semantyki i częstości użycia wyodrębnionych jednostek.

W przypadku korpusu języka czeskiego, a zwłaszcza korpusu synchronicznego SYN v8 uwzględnionego w badaniu, procedura jest znacznie uproszczona dzięki wbudowanemu narzędziu *FRANTA*, którego szczegółowy opis przedstawiliśmy w podrozdziale 3.8.4 *Kolokacje w Narodowym Korpusie Języka Czeskiego*. Dzięki wskazanej funkcji korpus udostępnia gotowe połączenia wielowyrazowe zawierające wybrany leksem, co znacząco skraca czas poszukiwań potocznych reпродукtów i pozwala ograniczyć analizę wyłącznie do prób określenia ich znaczenia semantycznego oraz częstości występowania wyekscerpowanych jednostek. Nasza analiza zostanie przeprowadzona w oparciu o leksem **VĚC**, którego artykuł hasłowy przedstawia się następująco:

věc ~i ž (1., 4. mn. ~i; 3. mn. ~em; 6. mn. ~ech; 7. mn. ~cmi) **1** konkrétní předmět nebo jev vnímaný smysly ▷ vzácná/starožitná/ztracená věc • Vaří samé dobré věci. **2** věci předmět denní potřeby ▢ majetek ▷ sbalit si své věci ▷ osobní věci **3** skutečnost; jev ▷ To je známá věc. • bojovat za dobrou věc = prosazovat, hájit něco prospěšného • nazývat věci pravými jmény = nebát se a umět říct kritický názor, mluvit bez příkras **4** událost ▷ Staly se divné věci. • To jsou mi věci! = zvláštní, divné události **5** námět, obsah jednání ▷ To je jeho věc. • To je věc názoru. ▷ věc cti/slušnosti • ministerstvo zahraničních věcí • jít přímo/rovnou k věci = začít se zabývat podstatou problému • mluvit k věci = neodbočovat od tématu • být od věci (poznámka ap.) = nesouviset s tématem • být nad věcí = mít potřebný nadhled, odstup od záležitosti • postavit koho před hotovou věc = sdělit někomu již rozhodnutou záležitost a tím ho zbavit možnosti podílet se na rozhodování • plést se/strkat nos do cizích věcí = zasahovat do cizích záležitostí **6** práce určená k vykonání • jít na věc = pustit se energicky do práce • umět vzít za věc = být zručný, pracovitý

Zawarta definicja w bogaty sposób ilustruje przykłady użycia jednostki w reprodukowalnych frazematyzacjach wielowyrazowych, wśród których znajdują się podejrzwane o stałą odtwarzalność połączenia, takie jak: *To je známá věc*; *Staly se divné věci*; *To jsou mi věci*; *To je věc názoru*. W kolejnym etapie odwołujemy się do funkcji *Kolokace* korpusu SYN V8 z zastosowaniem parametru *col_lemma* oraz zmienności opcji *V kontextu*, aby odnaleźć najczęściej używane kolokacje (bądź gotowe idiomy), które zostały zaprezentowane poniżej:

Hledat v korpusu

syn v8 ☆

Pokročilý dotaz ☐ | Klávesnice | Předchozí dotazy | Interpretace dotazu

věc

Kolokace

Atribut: col_lemma [collocation lemma]

V kontextu: -5 do 5

Minimální frekvence kolokátu v korpusu: 3

Minimální frekvence kolokátu v kontextu: 3

Kolokační míry:

Zobrazit Použít pro třídění

T-score	☒	☐
MI	☒	☐
MI3	☐	☐
log likelihood	☐	☐
min. sensitivity	☐	☐
logDice	☒	☒
MI.log_f	☐	☐
relative freq.	☐	☐

Vytvořit seznam kolokátů

2.	p / n	osobní_věc	22 614	10,16	150,25	8,26
5.	p / n	to_je_jeho_věc	7 436	9,17	86,08	6,66
6.	p / n	není_od_věci	7 131	9,57	84,33	6,61
8.	p / n	podstata_věci	4 776	10,15	69,05	6,04
14.	p / n	druhá_věc_je,_že	2 643	8,91	51,3	5,18
18.	p / n	jak_se_věci_mají	2 214	9,16	46,97	4,93
43.	p / n	věc_názoru	1 263	10,17	35,51	4,12
50.	p / n	to_je_jedna_věc	1 115	9,38	33,34	3,94

Jak možná zauvažyc, w niektórych prípadoch korpus wyodrębnił konstrukcje wielowyrazowe, które prawdopodobnie występują w poświadczeniach tekstowych jako samodzielne i gotowe reprodukt y komunikacyjne. W celu potwierdzenia postawionych hipotez i sprawdzenia otoczenia kontekstowego dla pozostałych wyników przeprowadzono weryfikację każdej kolokacji, której rezultaty przedstawiono poniżej:

osobní_věc	<i>To je osobní věc; Je to osobní věc každého</i>
to_je_jeho_věc	<i>Není to moje věc; To je vaše věc</i>
není_od_věci	<i>To je úplně od věci; To není od věci</i>
podstata_věci	<i>Podstatou věci je...; To je podstata věci</i>
druhá_věc_je,_že	Gotowy reprodukt.
jak_se_věci_mají	Gotowy reprodukt ze znakiem zapytania
věc_názoru	<i>To je věc názoru</i>
to_je_jedna_věc	Gotowy reprodukt.

Zgodnie z przypuszczeniami, trzy spośród kolokacji wskazanych przez korpus stanowiły gotową konstrukcję reproduktu wielowyrazowego, natomiast wyrażenie *To je věc názoru* było jedynym, które odpowiadało wyrażeniom potocznym zaproponowanym przez słownik. W przypadku pozostałych reproduktów z tabeli można zauważyć, że większość z nich charakteryzuje się znacznym podobieństwem strukturalnym, różniąc się jedynie obecnością odmiennych lub dodatkowych elementów, np. *To je osobní věc* - *Je to osobní věc každého*. Pomimo bezspornego podobieństwa strukturalnego, założenie o ich identycznym zastosowaniu w sytuacji komunikacyjnej byłoby niewłaściwe. W związku z tym niezbędne jest zwrócenie się do poświadczeń korpusowych w celu określenia różnic semantycznych. Wyniki analizy powyższych reproduktów z uwzględnieniem ich zastosowania w komunikacji, znaczenia i funkcji są następujące:

<i>To je osobní věc</i>
Komentarz reakcyjny (na początku zdania) Używany w sytuacji: 1) Gdy mówiący nie chce odpowiedzieć na pytanie rozmówcy, ponieważ podejmowany temat dotyczy jego spraw osobistych, których nie chce poruszać.
1) „ <i>O co jde? “ chce vědět a prohlíží si mě s podezřívavě povytaženým hustým tmavým obočím. Chápu, že Drakea nesnáší Ted. Přebíral mu holku. Ale co proti němu máš ty? “ „To je osobní věc, “řeknu.</i>

<i>To je osobní věc každého</i>
Komentarz wyjaśniający (w środku wypowiedzi)/podsumowujący (na końcu zdania). Używany w sytuacji: Dla podkreślenia akceptowalności ze strony mówcy faktu, że poruszana kwestia dotyczy indywidualnych wyborów każdego człowieka i nie należy ich oceniać.
1) <i>Znáť navzájom své silné a slabé stránky a počítat s nimi. Respektovat přání a zájmy toho druhého. To zní asi jako klišé, ale je to tak. Je papír na vztah důležitý? To je osobní věc každého, ale podle mě na papíru nezáleží.</i>
2) <i>Konečně někdo mluví o mých filmech, a ne o tom, jestli se rozvádím, nebo ne! To je strašně příjemné. “To nemá cenu rozebírat. Taky jsem si to prožil a to je osobní věc každého”.</i>

<i>Není to moje věc</i>
Komentarz podsumowujący. Używany w sytuacji: 1) Dla wyrażenia braku zainteresowania omawianą sprawą, zdystansowania wobec niej i niechęci oceniania pewnych kwestii 2) Jako

sugestia ukrytej opinii, którą mówiący chce przekazać, ale nie chce brać odpowiedzialności za swoje słowa.

1) Petr Nečas kritizoval i podporu Pussy Riot. I v tomto se s ním shodujete? - Nic o tom nevím. Nebyl jsem u toho, tak to nechci hodnotit. **Není to moje věc.** **2)** Důvodem protahování playoutu je zřejmě playoff první ligy. „To je možné, ale i ta se dala určitě udělat jinak. Já se k tomu už radši více vyjadřovat nebudu. **Není to moje věc**, ale svůj názor jsem si říci prostě musel.”

To je vaše věc

Komentář reakčný. Używany w sytuacji: 1) Dla podkreślenia, że decyzja w sprawie określonego działania zależy wyłącznie od działania rozmówcy. Wyrażenie zdystansowania wobec podejmowanego tematu.

1) Jestli náš výkon proti Jablonci ale vypadal jako krize, tak doufám, že krizi řeší všechny ostatní týmy v lize, “poznámenal kouč Pavel Vrba. „Možná je to na odvolání trenéra. **Ale to je vaše věc**, co napíšete”.

To je úplně od věci

Komentář oceniający. Używany w sytuacji: Dla podkreślenia, że wypowiedź rozmówcy nie ma związku z głównym tematem omawianego zagadnienia, i jest nieodpowiednia w podejmowanym kontekście rozmowy.

1) Přece nemáme žádné mořské draky, abysme je s něčím porovnávali. Můj drak je prostě podstatně menší než všichni ostatní. **To**, co říkáš, **je úplně od věci.**”

To není od věci

Komentář reakčný (na počátku zdania) . Używany w sytuacji: Dla zasygnalizowania, że pomysł bądź sugestia rozmówcy jest trafna w kontekście omawianej sprawy i jest warta rozważenia 2) Dla podkreślenia przydatności pomysłu, który zdaniem mówiącego ma swoje uzasadnienie w kontekście przedstawionej sprawy

1) Moje žena se mě snaží vykrmit, ale zuby a sliny i chuť chybějí, jde to těžko. Mám jen kašovitou stravu. “Vzal jste si třeba na pomoc psychologa?” „Asi **to není od věci**, ale já si psychologa nikdy nebral. **2)** Každý se bojí o práci. Sebeobranu do školních osnov poznámka může to vypadat jako pokus o vtip, ale určitě **to není od věci.**

<i>Podstatou věci je, že...</i>
Komentarz wprowadzający. Używany w sytuacji: Dla przejścia w rozmowie do najważniejszego aspektu omawianego zagadnienia i wyjaśnienia głównej istoty sprawy.
1) <i>Za deset let zvýšily zadlužení země ze zhruba 400 miliard na 1700 miliard. Je třeba jasně říci, že ústavu je nutno dodržovat. Spory o dodržování formy by však neměly zakrývat podstatu věci. Podstatou věci je, že naše země nevzkvétá, naopak.</i>

<i>To je podstata věci</i>
Komentarz podsumowujący. Używany w sytuacji: Dla wskazania na kluczowy i najważniejszy aspekt omawianej sprawy i skierowaniu uwagi rozmówcy na podstawową kwestię danego zagadnienia.
1) <i>Vaří se také z málo obvyklých surovin , je to trošku alchymie. I na soutěžích vidím , že už to často s kuchařinou nemá mít nic společného. Ale v našem oboru je důležité zaujmout - to je podstata věci.</i>

<i>Druhá věc je, že...</i>
Komentarz wprowadzająco-rozdzielający (w środku wypowiedzi). Używany w sytuacji: Dla wprowadzenia dodatkowego aspektu omawianego zagadnienia, który jest powiązany z poprzednią wypowiedzią mówiącego, i oddzielenia przekazanych informacji.
1) <i>A i kdyby nakonec nějaké měli, ztratí onu znalost a nadhled tím, že se pohybují příliš dlouho v rámci jednoho prostředí. To je jedna věc. Druhá věc je, že se mnoho zákazníků snaží, a je to dáno nejen rozpočty, udržet hlavně provoz.</i>

<i>Jak se věci mají?</i>
Komentarz wprowadzający. Używany w sytuacji: Dla sprawdzenia aktualnych okoliczności określonej kwestii, które może wyrażać się poprzez pytanie w kontekście konkretnej sprawy bądź ogólnego stanu rzeczy.
1) <i>Promiň, že mám volno až teď. Howard Hawks mě zdržel v Chicagu , no hrůza. To nevadí, strýčku Phile. Posad' se. Takže ...jak se věci mají? Rose říkala , že tu chceš bydlet.</i>

<i>To je věc názoru</i>
Komentarz podsumowující. Używany w sytuacji: 1) Dla podkreślenia różnorodności opinii w kwestii omawianej sprawy, które na swój sposób są subiektywnie uzasadnione 2) Dla neutralizacji treści swojej wypowiedzi, która mogłaby wywołać potencjalny konflikt z rozmówcą.
1) - <i>Co řeknete na názor, že pro naši společnost by bylo zdravé, aby ji příští prezident už nevedl za ruku - aby to byl obyčejný ouřada? Jaký má být prezident, to je věc názoru. 2)</i> <i>Ale já jsem ke svým kamarádům loajální, i když dělají chyby. Ode mě o žádném kamarádovi neuslyšíte nic špatného. LN: Ani kdyby této zemi škodil? Nemyslím, že by někdo z nich zemi škodil. Vy novináři této zemi škodíte mnohdy více. Ale to je věc názoru.</i>

<i>To je jedna věc</i>
Komentarz wprowadzająco-rozdzielający (w środku wypowiedzi). Używany w sytuacji: Aby oddzielić pewien aspekt sprawy, o której mówiący chce wspomnieć, jednocześnie wskazując na inną ważną kwestię, powiązaną z omawianym tematem.
1) <i>Ve smyslu bezprostředního zhodnocení a uplatnění je filozofie "nepotřebná" - to je jedna věc. Druhá věc je, že filozofie se nedá použít, že není k dispozici pro účely ležící mimo ni samu, že je sama účelem 2)</i> <i>Nejsou - li po ruce umírající lidé? „Zkoumání na zvířatech bylo neprůkazné. “To je jedna věc. „Můžu jen dodat , že badatel – mám na mysli sebe – musel být schopen subjekt instruovat.</i>

W danym przypadku wyekscerpowane czeskie reproduktory wielowyzrazowe, zawierające leksem *věc*, stanowiły głównie przykłady komentarzy wprowadzających i podsumowujących, a także gotowe repliki. Obserwacja list konkordancyjnych wygenerowanych przez korpus umożliwiła wstępnie założyć, że w przypadku reproduktu *To je vaše věc* można sformułować konstrukcję *To je <koho> věc*, w której elementem zmiennym są zaimki dzierżawcze, których wymiana nie wpływa na znaczenie reproduktu, a jedynie wskazuje na inną osobę (wyjątkiem w tej grupie byłby zaimek *moje*, który charakteryzuje się inną semantyką). Warto również zauważyć, że dwa z wyodrębnionych reproduktów stanowiły w wielu poświadczeniach przykłady jednostek, współwystępujących w tekście w bezpośredniej bliskości, co sugeruje ich potencjalną zależność funkcjonalną w analizowanym kontekście, tj. *To je jedna věc. Druhá věc je, že...* Końcowy etap badania zakładał oszacowanie stopnia frekwencji dla każdej wskazanej

jednostki wielowyrazowej w oparciu o zasoby korpusu SYN v8, Araneum Bohemicum Maximum oraz korpusu ORAL v1:

REPRODUKT	FREKWENCJA
<i>To je osobní věc</i>	65 poświadczeń
<i>Je to osobní věc každého</i>	27 poświadczeń
<i>Není to moje věc</i>	117 poświadczeń
<i>To je vaše věc</i>	403 poświadczenia
<i>To je úplně od věci</i>	2 poświadczenia
<i>To není od věci</i>	101 poświadczeń
<i>Podstatou věci je...</i>	92 poświadczenia
<i>To je podstata věci</i>	80 poświadczeń
<i>Druhá věc je, že...</i>	3522 poświadczenia
<i>Jak se věci mají?</i>	24 poświadczenia
<i>To je věc názoru</i>	840 poświadczeń
<i>To je jedna věc</i>	1469 poświadczeń

Przy zaprezentowaniu powyższych danych należy zauważyć fakt, iż niska frekwencja niektórych z wymienionych wielowyrazowców wynika z określenia liczby poświadczeń wyłącznie dla wskazanej formy tekstowej, a nie jej potencjalnych wariantów o identycznej semantyce i funkcji komunikacyjnej, których dodatkowe uwzględnienie zwiększyłoby liczbę poświadczeń (np. *To je osobní věc/Je to osobní věc/To osobní věc* – suma ich poświadczeń wynosi 415). Z kolei reprodunkt *To je úplně od věci* i jego zarejestrowanie w zaledwie dwóch konkordancjach skłania do odrzucenia przyporządkowania danej jednostki wielowyrazowej do odtwarzalnych wyrażen emotywno-reakcyjnych. Łączna liczba pozyskanych przykładów tekstowych dla wszystkich reprodunktów wyniosła najwięcej ze wszystkich analiz, bo aż 6 772 poświadczeń tekstowych, co świadczy o reprezentatywnych zasobach tekstowych NKJCz. Udział poszczególnych korpusów w przeprowadzonej analizie wyniósł procentowo:

1. SYN v8 – 61% (4 160 poświadczeń);
2. Araneum Bohemicum Maximum –38 % (2 569 poświadczeń);
3. ORAL v1 – niespełna 1% (43 poświadczenia).

Przedstawione dane liczbowe zdają się potwierdzać fakt, iż korpusy języka mówionego w dalszym stopniu wymagają wzbogacenia większą liczbą tekstów, a aktualne zasoby korpusu

ORAL v1 w liczbie 5,4 miliona słów okazują się niewystarczające podczas prób wyszukiwania wielowyrazowych jednostek języka mówionego. Zadanie rejestrowania żywej mowy przejmują w takim wypadku korpusy tekstów pisanych, które z jednej strony niewątpliwie zawierają w tekstach przykłady reproduktów języka mówionego, a z drugiej – nie powinny być głównym narzędziem podczas analiz tej odmiany języka. Poniżej przytoczono inne kolokacje o wysokiej frekwencji, które odnaleziono poprzez wskazanie w początkowej części analizy leksemu jednowyrazowego i dzięki temu uzyskano następujące reproduktów wielowyrazowe:

WYRAZ	POŁĄCZENIA	REPRODUKTY WIELOWYRAZOWE
ani	<i>ani_nemluvě; nemít_ani_tušení; někomu_to_ani_nepřijde</i>	<i>Ani nemluvě; Nemám ani tušení; Ani mi to nepřijde</i>
blbost	<i>co_je_to_za_blbost; kecat_blbosti; blbost_je_věčná</i>	<i>Co je to za blbost?; Nekecej blbosti; blbost je věčná</i>
chápat	<i>Co se děje; jak_se_to_mohlo_stát; moc_dobře</i>	<i>Nechápu, co se děje; Nechápu, jak se to mohlo stát; Chápu moc dobře</i>
kapitola	<i>To_je_jiná_kapitola; to_už_je_jiná_kapitola</i>	<i>To je jiná kapitola; To už je ale jiná kapitola</i>
náhoda	<i>to_je_náhoda; náhoda_je_blbec; nehoda_není_náhoda</i>	<i>To je ale náhoda; Náhoda je blbec; Nehoda není náhoda</i>
případ	<i>v_tomto_případě; v_případě_potřeby; v_nejlepším_případě</i>	<i>V tomto případě; V případě potřeby; V nejlepším případě</i>
říkat	<i>Co_tomu_říkáš; nikdy_neříkej_nikdy; tak_říkajíc</i>	<i>Co tomu říkáš?; Nikdy neříkej nikdy; Tak říkajíc</i>
snad	<i>to_snad_není_možné; to_snad_není_pravda</i>	<i>To snad není možné; To snad není pravda</i>
věřit	<i>věř_nevěř; nechtít_věřit; to_mi_věř</i>	<i>Věř nevěř; Nechci tomu věřit; To mi věř</i>
znít	<i>otázka_zní; zlaté_pravidlo; moc_dobře</i>	<i>Otázka zní...; Zlaté pravidlo zní...; Zní to moc dobře</i>

Zaprezentowana metoda słownikowo-kolokacyjna umożliwiła skuteczne odnalezienie, wyodrębnienie oraz identyfikację jednostek pod względem ich semantyki i pełnionej funkcji w wypowiedzi, a także umożliwiła weryfikację stopnia ich rozpowszechnienia. W przypadku analizy jednostek polskich *Wielki Słownik Języka Polskiego PAN* oraz wybrane polskie korpusy tekstowe okazały się użytecznymi narzędziami, dostarczając łącznie 2696 poświadczeń tekstowych. W perspektywie przyszłych badań należy wziąć pod uwagę występującą silną

dysproporcję w liczbie dostarczonych przykładów tekstowych przez poszczególne korpusy, spośród których najskuteczniejszym okazał się internetowy zbiór tekstowy MoncoPL. Podobnie w języku rosyjskim i czeskim, odpowiednie zasoby korpusowe i leksykograficzne pozwoliły na wyodrębnienie znaczącej liczby konkordancji tekstowych, wynoszących odpowiednio 2543 oraz 6772 poświadczenia, których dostarczyły przede wszystkim korpusy Araneum Russicum Maius i SYN v8. Przeprowadzona analiza ujawniła dwa istotne ograniczenia zastosowanej metody, polegające z jednej strony na ryzyku pominięcia niektórych reproduktów wskutek opierania się głównie na intuicji językowej badacza oraz subiektywnej selekcji kolokacji, a z drugiej – na nieuniknionej czasochłonności związanej z ręczną selekcją haseł słownikowych i późniejszych poświadczeń korpusowych. Kolejnym wyzwaniem okazała się różnorodność wariantów danej jednostki, które z uwagi na swą zmienność strukturalną nie poddają się łatwej identyfikacji, co ponownie wydłuża czas analizy i zwiększa ryzyko pominięcia istotnych reproduktów językowych. Ostatecznie badanie potwierdziło przydatność metody słownikowo-kolokacyjnej w wyodrębnianiu fatyzmów i emotywów, mających ugruntowane miejsce w codziennej komunikacji mówionej i pisanej. Niektóre z wykorzystanych korpusów okazały się zasobami o wysokiej reprezentatywności, co sprzyjało identyfikacji najczęściej używanych kolokacji. Wskazane wyniki sugerują jednak potrzebę zarówno dalszych badań, jak i wzbogacenia zasobów tekstowych korpusów (w szczególności języka mówionego), co pozwoliłoby na pełniejsze uchwycenie aktualnych tendencji w zakresie jednostek wielowyrazowych w analizowanych językach.

6.2 Pomocniki formalne

W celu nieautomatycznego dochodzenia do jednostek wielowyrazowych wykorzystane mogą zostać zaprezentowane poniżej metody naprowadzające, które wskazują na wielowyrazowe konstrukcje, mogące się charakteryzować regularną odtwarzalnością w języku:

1. Pomocniki formalne – Zgodnie z definicją Wojciecha Chlebdy przez pomocniki formalne należy rozumieć „Zawarte w tekście wypowiedzi znaki bądź sformułowania, których autor wypowiedzi użył jako świadectw swej wiedzy lub przeświadczenia o tym, że pewne fragmenty jego narracji, położone najczęściej w bezpośredniej bliskości tych wskaźników są ponadjednostkowym dobrem wspólnym” (Chlebda 2010b: 21). Wśród wymienionej grupy wyróżniony zostaje cudzysłów i tzw. metaoperatory ogólne i względne.

a) Cudzysłów służy dla zasygnalizowania, że użyte w tekście wypowiedzi sformułowanie jest niejako powszechnie znane wśród określonej grupy ludzi i zostało

zaczepnięte z innego źródła (pełniąc tym samym funkcję poświadczającą). Jednak sens tego, co cudzysłów graficznie oddziela każdorazowo wymaga szczegółowej interpretacji, wynikającej z wielofunkcyjności owego znaku. Podstawowe zastosowanie cudzysłowu sprowadza się do zacytowania wypowiedzi jednostki, a nie wspólnego odtwarzalnego kodu wśród określonej wspólnoty językowej. Dlatego następujące przykłady użycia cudzysłowu (i tego, co mieści się w granicach znaku interpunkcyjnego) w ramach badań nad reprodukowalnymi jednostkami wielowyrazowymi nie będą rozpatrywane (przykłady z NKJP, NKJR i InterCorp v16ud):

W Wykładach z filozofii dziejów Hegel dotyka postaci Jezusa...Nie ma oczywiście mowy o zaprzeczeniu Jego istnienia, niemniej: „Niesłusznie postępujemy również, gdy mówimy o Chrystusie tylko jako o historycznej postaci, która kiedyś istniała. Pytamy wówczas: jakże było z Jego urodzeniem?”

Например, довольно влиятельный митрополит Воронежский и Липецкий Мефодий (Немцов), правда, по поводу региональных выборов говорил: "Если нам известно, что тот или иной политический деятель настроен резко против Церкви Христовой, то мы должны сказать об этом верующим".

Věci, které nás zastihnou nepřipravené na naši cestě životem. Svátý František (z Assisi) kdysi prohlásil: "Pane, dej mi trpělivost ,abych snesl to, co se změnit nedá".

Jednak wyodrębnienie kolejnych konkordancji, zawierających cudzysłów i znajdujący się pomiędzy jego granicami ciąg wyrazów, pozwala na wstępne założenie, że wyodrębnione jednostki charakteryzują się frazematycznością:

Rodzice zaciągają pożyczki. Wynajmują mieszkania, a sami przenoszą się do dziadków. Rezygnują z wakacji, książek, kina. Szukają dodatkowej pracy. Proszą firmy o sponsorowanie studiów córki czy syna (na ogół bezskutecznie). Naruszają oszczędności „na czarną godzinę”.

Вроде бы публика привыкла к новому уровню цен и ощутимого закрытия печатных СМИ не происходит, вместе с тем желающих платить немалые деньги за "кома в мешке" становится все меньше

Ačkolí musí rozpočet vyčleněný na zemědělství zohledňovat nepředvídatelné situace i v případě přebytku, mám několik námitek k tomu, aby byl neustále používán jako "zlatý důl", který slouží na úhradu nepředvídatelných událostí.

W danych przypadkach znak cudzysłowu mógłby zostać zamieniony wyrażeniami, takimi jak *(na) tak zwany/ как говорят/ příslovečný*. Z punktu widzenia stopnia wiarygodności wyodrębnionej jednostki wielowyrazowej, najbardziej korzystną sytuacją jest jednocześnie i bliskie występowanie obok siebie w tekście znaku cudzysłowu i operatora, które zachowują swoją funkcję testimonialną⁹, np.

<i>Jest nieufna i ostrożna, a tego Marcie zawsze brakowało w kontaktach z ludźmi. Do każdego podchodziła z zaufaniem, życzliwością i otwartym sercem, <u>jak to się mówi "na dłoni"</u>. Teraz zbiera plony takiej naiwności.</i>
<i>—А сунешься куда покруче лжеменеджера, так: «Ваш возраст?.. Такие нам не требуются...» <u>Как говорится, «моя года — мое богатство».</u></i>
<i>Avšak ještě než se podíváme na důsledky, mi dovolu, abych vám představil dalšího neklidného ducha: osiku, jejíž <u>príslovečné chvění</u> - „<u>Třese se jako osika</u>“.</i>

b) W ramach wspomnianych metaoperatorów można wyodrębnić metaoperatory bezwzględne i względne.

Pierwsze z nich zawierają informację o tym, że pewien ciąg wyrazowy jest regularnie odtwarzany przez ogół użytkowników danego języka bez uwzględnienia innych kryteriów, np. *przysłowiowy; tak zwany; jak to mawiają; jak to mówią; как говорится; как принято говорить; что называется; jak se říká; jak se praví; jak by se tak řeklo* itp.

Poniżej przytoczone zostały przykłady konkordancji, zawierające wspomniany rodzaj operatorów, które traktować należy jako swoistą „nawigację” lub odpowiedź, sugerującą obecność jednostki wielowyrazowej:

<i>Kościół zawdzięcza swoją pozycję walce, inkwizycji, prześladowaniem innowierców, tylko przemoc w słusznej sprawie przynosi efekty, <u>jak to się mówi "cel uświęca środki"</u>.</i>
<i>Przepraszam, za wyjątkiem komisji ds. osób niepełnosprawnych. Reszta przepadła <u>jak przysłowiowy "kamień w wodę"</u>.</i>
<i>Pokazuje ona bowiem dowodnie, <u>jak bezsilna jest we współczesnym świecie tak zwana czysta logika, tak zwany zdrowy rozsądek</u>.</i>
<i>Dawno, dawno temu, <u>jak to się zwykło mówić "za komuny"</u> był taki dowcip, że w socjalizmie są właściwe tylko cztery kataklizmy: wiosna, lato, jesień i zima.</i>

⁹ Od łac. *testimonium* ‘świadectwo’. Termin A. M. Lewickiego, który pisał: „Przy stereotypach intencjonalnie ekspresywnych pojawia się często operator testimonialności (wskazania źródła wyrażenia) typu *jak to mówią, przysłowiowy*” (Lewicki 2003: 31). W. Chlebda uznaje to z kolei za przejaw dawania świadectwa typowości i powtarzalności danego wyrażenia w określonych sytuacjach (Chlebda 2010b: 23).

<i>Сегодня мы слушали человека, который играет ключевую роль в корабельной ядерной энергетике, так что получили, как говорится, <u>информацию из первых рук</u>.</i>
<i>Но я понимал, что у меня может возникнуть так называемый «<u>стокгольмский синдром</u>» — чувство единения с теми, с кем рядом находишься под пулями.</i>
<i>«Наши отношения с другими странами должны быть также нацелены на решение задач модернизации России. Нам нечего, как принято говорить, «<u>надувать щёки</u>».</i>
<i>— Если орган здоров, должны преобладать положительные эмоции? — Да. Недаром говорится: «<u>в здоровом теле — здоровый дух</u>».</i>
<i>Domnívali se, že každý člověk je celý svět v miniatuře, jakýsi „mikrokosmos“, který je odrazem „makrokosmu“. To vedlo k myšlence, že existuje všeobecně platné právo, tak zvané „<u>přirozené právo</u>“.</i>
<i>Byl jsem sám jako ten přísluvěčný <u>kůl v plotě</u>. Nikdo se mnou nebyl. Vůbec nikdo. Nikdy.</i>
<i>Neumím si dost dobře představit, že bych vyměnil post moderátora za práci v terénu, i kdybych si mohl sám vybírat témata reportáží. Víš, jak se to říká: <u>znám tu práci jako své boty</u>.</i>
<i>Od ledovců Ala-Mongju zavál chlad - a ty prastaré ledovce byly docela blízko, jak se říká, <u>na dosah ruky</u>.</i>

W przytoczonych poświadczeniach ponownie unaocznia się tendencja do jednoczesnego występowania w tekstach metaoperatora bezwzględnego wraz ze znakiem cudzysłowu, potwierdzającego status rozpowszechnienia jednostki wielowyrazowej w języku. Z kolei, metaoperatory względne rozpatrywane są w odniesieniu do konkretnych płaszczyzn rzeczywistości:

- miejsca (np. *jak mówią w Rosji; как говорят в США; jak se říká v Itálii*),
 - czasu (np. *jak to się kiedyś mówiło; как прежде говорили; jak se dnes říká*) i
 - wspólnoty (np. *jak to mówią politycy; как говорят арестанты; jak říkají hudebníci*).
- Pierwsze niosą informację o zwyczajach nazewniczych w określonym kraju czy regionie, które były bądź są obecnie używane w języku, np.:

Stworzyły sytuację, delikatnie mówiąc, patową. Pierwsza zrozumiała to Wanda: – No to – jak mówią w Polsce – <u>na drugą nóżkę</u> .
W połowie 1998 r. rząd przyjął program restrukturyzacji hutnictwa (miał kosztować 9 mld zł) - we wrześniu 2000 r. było już " <u>po ptokach</u> ", jak mówią na Śląsku , i rząd zaczął pracować nad nowymi koncepcjami.
- Przecież wojna się już skończyła, miasto patrolują jednostki międzynarodowej policji, a bośniaccy policjanci wyglądają na uprzejmych - oponuję. - Tak, ale, wie pani, jak tu się mówi - <u>noc nie ma świadków</u> ...
Opowieści, że Samoobrona walczy o interesy ludzi pracy to „ <u>zwykła popelina</u> ” jak mówią w Warszawie .
Я был для него, как говорят в Грузии , « <u>гость от Бога</u> » — оказался первым посетителем и понравился хозяину образом подгулявшего богатого господина, по традиции навестившего его заведение рано утром.
Они не вызывают подозрения только тогда, когда несут дрова: небольшое бревнышко или, как здесь говорят , " <u>палку дров</u> " на плече.
Тогда ребенок и в себе может либо иметь структурную иерархию, как говорили на Руси — « <u>Царя в голове</u> », тогда все его низшие силы души будут управляться высшими
Давно я Вас не видел. Ну, а пока желаю жить-поживать, и Вам того же и от всей души. И так, чтоб, как говорят у нас в Архангельском , <u>и голова была на плечах, и вода в ручьях</u> .
Zastav! Dopravím tě do bezpečí, ať chceš nebo ne. (<u>chca nechca jak říkáme na Moravě</u>). Máš zastavit!
Jeho vztahy s Kongresem zdaleka nejsou hladké. Obama ho však nyní potřebuje víc než kdykoliv dříve – aby mu, jak říkají v Americe , „ <u>kryl zadek</u> “.
Z toho důvodu můžeme udělit osvobození od HACCP a sledování hygieny Unii jen v těch případech, kdy jsme si naprosto jistí, že tím nezpůsobíme žádnou škodu. Jak se říká v Mad'arsku , je lepší se strachovat než se jednou zděsit.
<u>Poslechněme si tvou verzi skutečnosti</u> . Jak tomu říkají v Rusku , pravda?

Ponadto obserwacja struktury danego rodzaju metaoperatorów umożliwiała opracowanie następujących propozycji konstrukcji, których wpisanie w wyszukiwarke ułatwia proces ekskserpcji wskaźników miejsca i sąsiadujących obok reproduktów, np. *jak mówią w/na...*; *jak mawiano w/na...*; *jak się mówiło w/na...*; *как говорят в/на...*; *как говорили в/на...*; *jak se říká*

v/na...; *jak říkají v/na...* itd. Poniżej przedstawiono listę metaoperatorów miejsca o najwyższej frekwencji, w których bezpośredniej bliskości znajdowały się reprodukowalne jednostki wielowyrazowe:

METAOPERATOR MIEJSCA	PRZYKŁAD
<i>Jak mówią w/na...</i>	<i>Wbrew opiniom, jakie znalazłem w tajnej teczce personalnej von Klugego (założonej w osobistej kancelarii SS-Reichsführera), feldmarszałek nie miał inteligencko-szlacheckich cech mięczaka. Szedł "w zaparte" na całego, jak mówią w Polsce.</i>
<i>Jak się mówi w/na...</i>	<i>Sam (ze względu na swoją narodowość) prezydentem być nie może, ale – jak się mówi w Kabulu - uzna za prezydenta nawet Kaczora Donalda, jeśli ten pozwoli mu tylko pozostać Emirem Północy.</i>
<i>Jak to się mówi w/na...</i>	<i>Na pewno Finkelsztejn to człowiek z rozległymi koneksjami, z pierwszorzędną – jak to się mówi w Rosji - kryszką, bez niej Zencowa nie wyciągnąłby z pewnej organizacji nigdy, za żadne pieniądze.</i>
<i>Как говорят в/на...</i>	<i>Впрочем, со словами всегда так... Он поглядел поверх моей головы. — Устраивайся поудобней, разговор у нас долгий. Как говорят в России, в ногах правды нет...</i>
<i>Как <где> говорят...</i>	<i>В поселении в бухте Абердин проживает около сорока тысяч человек те, кто, как здесь говорят, принадлежит к «расе лодок» люди народностей танка и хокло.</i>
<i>Jak se to říká v/na...</i>	<i>Ale kde a kdy ji najít? Nakonec jsem si začal myslet, že pravá láska pro mě není. Jak se to říká v Polsku? Klobása není pro psy.</i>
<i>Jak říkají v/na...</i>	<i>Zklamal jsem tě někdy? Je od tebe velmi milé, že se chceš taky přiznat, Jing, ale jak říkají v Číně - Jedno dítě postačí.</i>
<i>Jak se říká v/na...</i>	<i>Ale to je, jak víte, pole příliš širé. Nejlíp když si člověk zvykne a stěžuje si jen střídmě. Koneckonců, jak se říká na Západě, zbývá toho hodně udělat: Chopme se toho!</i>

W danym miejscu warto podkreślić, że za najczęstszy rosyjski wskaźnik tekstowy należy uznać konstrukcję *как говорится в/на...*, której analiza poświadczeń wykazała jej zastosowanie w odniesieniu do miejsca zapisu konkretnych sentencji, anegdot czy przejęciu funkcji metaoperatora bezwzględne, np. *И тогда, как говорится в Евангелии,*

восплачутся все племена земные (Мф. 24: 30), каждый человек станет перед лицом этого последнего Суда.; Но, **как говорится в анекдоте**, неприятный осадок остался.

Poczynione obserwacje pozwalają wysnuć wniosek, iż strukturę **как говорится в/на...** wyjątkowo rzadko stosuje się w rosyjskich źródłach tekstowych w funkcji metaoperatora miejsca, dla którego w zasobach NKJR znaleziono tylko jeden przykład użycia: *Выгнать из училища, сделать сургуна, как говорится в Турции, то есть изгнать из страны; заморить даже в тюрьме, все это можно было сделать с учителем, своим подданным.*

Kolejną grupę operatorów względnych stanowią wskaźniki temporalne, odsyłające odbiorcę do praktyk nazewniczych w konkretnym przedziale czasowym, np. *jak się mówiło za PRL-u; jak mawiano w starożytności; mówiąc językiem dzisiejszym; как прежде говорили; как говорили в <какие> годы; jak se dřív říkalo; jak říkávali <kdy> itp.:*

<i>Głównym bohaterem, podmiotem dzieła, stał się pojedynczy człowiek, "<u>wyindywidualizowany z rozentuzjasmowanego tłumu</u>" – jak to się kiedyś mówiło.</i>
<i>Jak wynika z warszawskiej prasy, różni przedsiębiorczy ludzie podejmowali kilkakrotnie starania, by otworzyć w mieście "kantor stręczenia małżeństw", czyli mówiąc językiem dzisiejszym - biuro matrymonialne.</i>
<i>Po herbacie, jak się mówiło w moich czasach. Po jabłkach, jak się mówi teraz.</i>
<i>Mieczysław Jałowiecki nie był, jak byśmy dziś powiedzieli, "ślepy antysemita". Jak każdy antysemita, na swojej drodze spotkał nawet przyzwoitych Żydów.</i>
<i>Николай Михайлович Карамзин, потомок татарских ханов; Николай Васильевич Гоголь с его украинскими, или, как прежде говорили, малоросскими корнями...</i>
<i>Он сам решительно отказался жить с матерью и переселился к друзьям в барак, где, как говорили в те годы, «жилье, барахло и шамовка» были общими.</i>
<i>Не просто подавил их, а отвернулся от всех эмоциональных и смысловых потоков, протекавших через него — вышел, как бы сказали сегодня, в полный офф-лайн.</i>
<i>Одно дело мать, которая меняет любовников как перчатки, а другое — незамужняя старшая сестра, которая устраивает свою личную жизнь. Как сейчас говорят — в активном поиске. Майя — несчастный ребенок.</i>
<i>A pokud jde o mně... já bohužel nejsem takový troufalec. <u>Nejsem žádný Danton, jak se říkalo za našich studentských časů.</u></i>
<i>Vážení a milí, konečně jsme se dočkali teplých dní, sluníčko se snaží seč může a prázdniny jsou na dohled. Mnozí z vás se v tomto čase už jistě vydávají na cesty, jak se dřív říkalo „za hranice všedních dní“.</i>

*Volba na první tábor v oblasti padla na Borová Lada. Ne tedy Borovou Ladu, **jak bylo zvykem říkat**, ale roztroušená dědina se skutečně jmenovala ta Borová Lada.*

*Ale bude si muset osvěžit paměť ohledně matematické dimenze jeho díla. Kde ty informace vyhrabat? V Národní knihovně? VNK, **jak říkali zamlada**.*

Ze względu na różnorodność składniową danych metaoperatorów, niemożliwym jest ustalenie analogicznej do metaoperatorów miejsca konstrukcji, która pozwalałaby na automatyczne wyzyskanie początkowej części wskaźnika tekstowego wraz z sąsiadującym obok reproduktem. Niemniej jednak analiza dostępnych poświadczeń tekstowych pozwoliła na wyodrębnienie przykładowego spisu metaoperatorów temporalnych przedstawionych w poniższym wykazie:

<i>Jak się mówiło w latach... (np. siedemdziesiątych)</i>	<i>Mówiąc językiem... (np. dzisiaj)</i>
<i>Jak mawiano za... (np. PRL-u)</i>	<i>Jak byśmy dziś powiedzieli....</i>
<i>Co określano mianem...</i>	<i>Jak to się kiedyś mówiło...</i>
<i>Mówiło się kiedyś, że...</i>	<i>Jak kiedyś mawiano...</i>
<i>Как говорилось.... напр. (в старину)</i>	<i><то>, что называлось...<чем?></i>
<i>Как сейчас говорят...</i>	<i>Как прежде говорили</i>
<i>Как бы сказали сегодня...</i>	<i>Как говорили... напр. (в советские годы)</i>
<i>Jak se říkalo... (np. v minulosti)</i>	<i>Jak bylo zvykem říkat...</i>
<i>Jak říkali... (np. dřív)</i>	<i>Jak by se dneska řeklo...</i>
<i>Jak bychom dnes řekli...</i>	<i>Kdysi se tomu říkalo...</i>

Wskaźniki wspólnotowe wiążą występujący ciąg wyrazów z określonym środowiskiem, grupą zawodową, członkami stowarzyszeń itp., wśród których panują inne zwyczaje nazewnicze poprzez stworzenie własnego kodu językowego, zrozumiałego tylko dla członków określonej wspólnoty, np. *jak mówią nauczyciele; mówiąc językiem więźniów; jak mawiają lekarze; как говорят няни; как это называют социологи; как именуют её анашисты; jak říkají lékaři; školsky řečeno; máme-li se vyjádřit po námořnicku* itd. Dany metaoperator może posłużyć do wyodrębnienia tych jednostek, które wchodzą w zasoby słownictwa konkretnych socjolektów czy profesjolektów:

<i>Ale z każdą minutą <u>uchodziło z nich, jak mówią sportowcy, powietrze</u>. Nic przeto dziwnego, że krakowianie zdołali uzyskać zdecydowaną przewagę.</i>
<i>On zaś był dla niej problemem, ale nie zagadką, gdyż był obcym, zatem, <u>mówiąc językiem policyjnym</u>, był "z urzędu podejrzany".</i>
<i>Ja nie przyszedłem tu, żeby Wam opowiadać koszałki-opałki, chciałem żeby było przynajmniej po połowie. <u>Fifti-fifti, jak mówią studenci</u>.</i>
<i>Мобильное рабство отменяется Но до весны оператора лучше не менять Вы пользуетесь мобильным номером уже пол года или больше? Значит, <u>как говорят социологи, вы в зависимости от сотового оператора</u>.</i>
<i>Поэтому Б и отвечает Как это нет, имея в виду, что нож есть. Представьте себе, что Б отвечает, <u>как это называют учителя в школе, «полным ответом»</u>.</i>
<i>Все ограниченные вещи, и не тожко самые существа, но также определенные сих содержания по времени и пространству; и потому вселенная, <u>как определяют новейшие философы, есть чин всех вещей изменяющихся</u>.</i>
<i>Podobá se skladbě, kterou <u>hudebníci nazývají: téma s variacemi</u>.</i>
<i>Všechny ty <u>nepoužívané svaly</u> , <u>jak říkají hoši</u>, co patří do volné váhové kategorie.</i>
<i>Ted' vyslechni ty mě. Tvoje idea o novém Bohu, to je abstraktní teorie, i když svým způsobem <u>neobyčejně aktuální, mám-li se vyjádřit jazykem našich intelektuálů</u>.</i>

Na podstawie poprzednich konstrukcji formuł wskaźnikowych można zauważyć, że również wśród metaoperatorów wspólnotowych przeważa struktura, uwzględniająca zwrot *jak mówią/ как говорят/ jak říkají* (kto?kmo?kdo?). Do najczęstszych postaci wskaźnikowych formuł wspólnotowych można odnieść poniższe konstrukcje, które ilustrują poświadczenia, wyodrębnione z zasobów korpusowych:

METAOOPERATOR WSPÓLNOTOWY	PRZYKŁAD
<i>Jak mówią... (kto?)</i>	<i>Nawet redaktorzy gazety „Dla dzieci” mieli z tym przedmiotem kłopoty. Ale <u>jak mówią</u> nauczyciele: <u>wystarczy się systematycznie uczyć i wszystko stanie się łatwiejsze</u>.</i>
<i>Mówiąc językiem (czyim? jakim?)</i>	<i>To jest, <u>mówiąc językiem</u> prawniczym, tzw. <u>zamiar ewentualny, czyniący Twoją winę umyślną</u>.</i>

<i>Мówiąc naszym (czym?)</i>	<i>I że ja w jego sprawie będę musiał podpisywać dokumenty, które uruchomią, <u>mówiąc naszym żargonem, procedurę rozpracowania.</u></i>
<i>Как говорят... (кто?)</i>	<i>Мы знаем, что в глаголах 2-го спряжения происходит, <u>как говорят специалисты,</u> <u>массовый сдвиг акцентной парадигмы.</u></i>
<i>Как это называют... (кто?)</i>	<i>Между тем именно они создают «<u>групповое сцепление</u>», <u>как это называют социологи.</u></i>
<i>Как выражаются... (кто?)</i>	<i>Не нужно искать редкую информацию или создавать неожиданные смыслы. Важен просто сам факт общения. Как выражаются лингвисты, фатическая функция.</i>
<i>Jak říkají... (kdo?)</i>	<i>Pár strážců stálo na tribunách a u vchodů, ale tady se žádný nepohyboval. Jurij byl, jak říkají sportovci, šťastně v cíli.</i>
<i>Řečeno... (jak?)</i>	<i>Řečeno obchodnický, <u>rakve byly v kursu</u> - a v průběhu několika týdnů si Oliver osvojil začnou míru zkušeností.</i>
<i>(co?) (kdo?) nazývají</i>	<i>Naproti tomu reálné gravitony jsou totožné s tím, co klasičtí fyzikové nazývají <u>gravitačními vlnami.</u></i>

Wśród danej podgrupy należy poświęcić szczególną uwagę na zjawisko wskaźników autorstwa jednostkowego, które reprezentują zbiorowość eptonimów i poprzez zapożyczenie u innego autora stały się dobrem wspólnym ogółu. Wypowiedzenia jednostkowe (tzw. *skrzydlate wyrażenia*), pomimo ich niezaprzeczalnego rozpowszechnienia u rodzimych użytkowników poszczególnych języków, bardzo rzadko są opatrzone dodatkowym wskaźnikiem tekstowym. Ponadto, jak wykazała przeprowadzona analiza, w danej grupie nie istnieją określone modele stosowane do przytaczania publicznie znanych zwrotów, a ich różnorodność potwierdzają przedstawione poniżej wyekscerpowane poświadczenia:

<i>Kiedys artystka śpiewała hity w stylu "Oczy cziornyje", dziś pozostaje raczej zanucić stary klasyk. <u>Ciemność widzę, ciemność?</u></i>
<i>Plakat ku czci wspaniałej postaci Pani Wolańskiej z filmu „Kogel-mogel”. Prawdziwa gratka dla fanów kultowego już tekstu: <u>Marian! Tu jest jakby luksusowo.</u></i>
<i>Często wyśmiewa się wałęsowskie "<u>jestem za, a nawet przeciw</u>". Jednak może się okazać, że w sieciowym świecie przyszłości zwycięży wielowartościowy model podejmowania decyzji.</i>
<i>Она, по-моему, не менее популярна, чем знаменитое черномырдинское: <u>«Хотели как лучшие, а получилось как всегда».</u></i>
<i>На слух звучало как «Плохой Форд», но грамотные могли прочесть на радиаторе слово «Bedford» — стало быть, не «плохой», а «постельный». Та еще постелька. Впрочем, как говорил классик, <u>«кому и кобыла — невеста».</u></i>
<i>Неровен час, у читателя после всех «Великих афер» сложится превратное впечатление, что волиебная формула (<u>«Украл, выпил, в тюрьму!»</u>).</i>
<i>Člověk v tísnì, připomínala vtipnou scénu z jedné klasické české filmové pohádky: „<u>Odvolávám, co jsem odvolal, a slibuji, co jsem slíbil</u>“. Zákon putoval jako horký brambor mezi poslanci.</i>
<i>Veškeré předchozí snahy ztroskotaly zejména na neochotě reformních komunistů, kteří, povzbuzeni Pithartovým "<u>možná slyším trávu růst</u>", úspěšně blokovali ve vládě i parlamentu jakékoliv pokusy o razantnější zpytování minulosti.</i>
<i>K této ostudě nejen ČSSD a Děčína, ale i všech občanů, kteří se neozvou, mohu jen parafrázovat slova klasika: "<u>Je něco shnilého ve státě českém</u>".</i>

Podczas próby wyodrębnienia wskazanych jednostek wielowyrazowych udało się ustanowić powielany w każdym z języków schemat metaoperatorów, które pomimo bardzo niskiej frekwencji występowania, odsyłały czytelnika do sformułowań zapożyczonych od innego autora, tj. 1. Zastosowanie słowa ‚klasyk‘ wraz z czasownikiem i dodatkowym elementem, np. *zanucić słowa klasyka; как говорил классик; parafrázovat slova klasika*, itp., 2. Zastosowanie przymiotnika dzierżawczego, wskazującego na autora przytoczonych słów, np. *wałęsowski; черномырдинский; Pithartový*.

W grupie metaoperatorów względnych znajdują się również operatory gatunkowe, które informują odbiorcę o tym, że zarejestrowana jednostka wielowyrazowa posiada status przysłowia czy maksymy i naturalnie tym samym poświadcza jej reprodukowalność, np. *jak mówi polskie przysłowie; jak głosi stara maksyma; ludowe powiedzenie mówi; мудрость*

Do tej grupy zaliczyć możemy również **operatory quasi-gatunkowe**, pod którymi należy rozumieć frazy i sformułowania w społeczeństwie powszechnie znane, niosące jednocześnie znamiona typowości i weryfikowane przez świadomość jako swojego rodzaju „hasło rozpoznawcze”, np. *zasada; formula; hasło; лозунг; принцип; девиз; fráze; heslo; kliše*. Charakterystycznym dla danego typu operatorów tekstowych jest nierzadkie współwystępowanie wskaźników powszechności, takich jak: *znany; obiegowy; распространённый; известный; typický; památný* itp.:

<i>Cieszę się, że w ich programach nie dominuje polityka! Przez wiele lat, zawsze do moich programów zapraszam kogoś zgodnie z hasłem: "Idzie nowe".</i>
<i>- Potwierdza się stara zasada, że <u>punkt widzenia zależy od punktu siedzenia</u>. Wcześniej było wiele deklaracji socjalnych, propracowniczych. Tymczasem nie czuje się jakiejś propracowniczej polityki PiS.</i>
<i>To wszystko brzmi prawie tak samo, ale jak głosi znany slogan reklamowy, "<u>prawie</u>", czyni <u>wielką różnicę</u>.</i>
<i>Я знаю, что на каждую контратаку Британии и любого другого ее союзника, который решит к ней присоединиться, Россия ответит по принципу «око за око».</i>
<i>Растет деловая активность, появляются новые фирмы с многомиллионными оборотами, для которых известный девиз «<u>время — деньги</u>» постепенно обретает практический смысл.</i>
<i>[...] решил выпустить новогодний номер с большим портретом Сталина и его нетленными словами под клише: «<u>Жить стало лучше, жить стало веселей</u>».</i>
<i>Tam také zazněla z úst polského ministra památná věta: <u>Nebojím se tolik německé síly, jako se začínám bát německé nečinnosti</u>.</i>
<i>Podle zásady, že <u>možné je všechno</u>, může mít svou cenu poukaz na to, že by se Marxovo proroctví mohlo snadno vyplnit.</i>
<i>Během prvních pěti let inzerovali hlavní podniky maloobchodu s potravinami pomocí hesel typu: <u>Nízké ceny každý den</u>.</i>

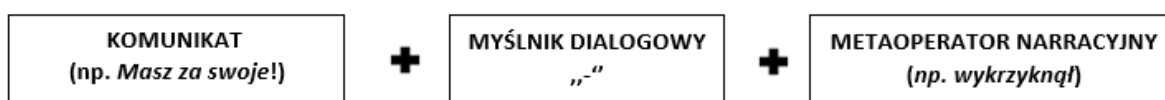
Wykorzystanie przedstawionych wskaźników formalnych niewątpliwie ułatwia i przyspiesza proces identyfikacji konstrukcji wielowyrazowych, które są reprodukowalne i zrozumiałe dla określonej wspólnoty językowej. Jedną z głównych zalet niniejszej metody jest możliwość systematycznego eksцерpowania reproduktów z korpusów językowych przy wykorzystaniu szerokiego repertuaru metaoperatorów. Dzięki temu sam proces badawczy staje

się bardziej obiektywny, ogranicza subiektywne interpretacje językoznawcy. Ponadto, analiza kontekstowa wskaźników dostarcza dodatkowych informacji o funkcjach językowych i cechach kulturowych badanych jednostek, co wzbogaca ich definicję semantyczną. Należy również wziąć pod uwagę fakt, że nie wszystkie konstrukcje wielowyrazowe w tekstach są wyraźnie oznaczone wskaźnikami tekstowymi, co może prowadzić do pomijania istotnych danych. Ponadto, różnorodność zastosowań cudzysłówów oraz metaoperatorów może powodować trudności interpretacyjne, szczególnie w przypadku ich wielofunkcyjności i wieloznaczności. Kolejną wadą jest potencjalna zależność wyników od jakości i wielkości analizowanego korpusu tekstowego, co znacząco ogranicza możliwość generalizacji wniosków. Zaprezentowana metoda wymaga niejednokrotnie uzupełnienia innymi metodami dochodzenia do jednostek wielowyrazowych w celu uzyskania nie tylko reprezentatywnej bazy reprodukowalnych jednostek fatycznych i emotywnych, ale także uwzględnienia wszystkich znaczeń semantyczno-pragmatycznych konkretnego wielowyrazowca.

6.3 Metaoperatory narracyjne

Polskie metaoperatory narracyjne

Źródłem zarówno metaoperatorów, jak i reproduktów pozostają teksty. Bogactwo dostępnej literatury, zarówno cyfrowej jak i papierowej umożliwia kontynuację procesu poszukiwawczego formuł wskaźnikowych i jednostek wielowyrazowych poprzez wnikliwą analizę samej struktury tekstu. Obserwacja zawartych w literaturze beletrystycznej wypowiedzi postaci, które należy traktować jako nieodłączny element świata przedstawionego określonej fabuły pozwala zauważyć, że wtrącenia narratorskie, typu *powiedział*; *powtórzył*; *westchnął* można rozpatrywać w kategorii *metaoperatorów narracyjnych*. Podstawowy schemat przytoczenia i opisanie komunikatu postaci literackiej przedstawia się następująco:



Przedstawiona kompozycja budowy komunikatu umożliwia więc opracowanie metody, polegającej na wykorzystaniu metaoperatorów narracyjnych w celu ręcznego wyodrębnienia reproduktów języka mówionego. Jednak procedura ta musi zostać poprzedzona samym odnalezieniem formuł wskaźnikowych. Umożliwia nam to wykorzystanie znaku myślnika dialogowego „-” (rozpatrywanego również jako rodzaj operatora), którego wpisanie w

wyszukiwarkę NKJP przynosi aż półtora miliona konkordancji, wymagających każdorazowej analizy struktury i treści, które znajdują się w bezpośredniej bliskości zadanego w zapytaniu znaku. Dla celów niniejszego badania wykorzystano tysiąc poświadczeń, z których wyizolowane zostały metaoperatory narracyjne przedstawione w poniższej tabeli:

FORMA WYJŚCIOWA	METAOPERATOR NARRACYJNY
Bełkotać	Bełkocze/Wybełkotał
Dodać	Dodałem/Dodał
Mówić	Mówię/Mówi
Myśleć	Myślę/Pomyślałem/Pomyślałam/Pomyślał
Powiedzieć	Powiedziałem/Powiedziałam/Powiedział
Powtórzyć	Powtarzam/Powtórzył/Powtórzyła
Rzucić	Rzuciłem/Rzuciłam/Rzucił
Upewnić się	Upewniłem się/Upewniłam się
Westchnąć	Westchnąłem/Westchnął/Westchnęła
Zainteresować się	Zainteresował się/Zainteresowała się
Zapewnić	Zapewniłem/Zapewnił
Zawołać	Zawołałem/Zawołał

Wyodrębniony spis trzydziestu metaoperatorów ilustruje wyłącznie te konstrukcje, które zostały wyszukane w wyselekcjonowanych przez NKJP poświadczeniach. Jednak sama ich obserwacja pozwala stwierdzić, że istnieje duże prawdopodobieństwo prawidłowego rozbudowania zakresu (występującego w literaturze) poprzez odmianę formy wyjściowej (lub pochodnej) przez odpowiednie kategorie gramatyczne: osoby, czasu i aspektu, np.:

Krzyczę
Krzyczałem/Krzyczałam/Krzyczał/Krzyczała/Krzyczało/Krzyczeliśmy/Krzyczeli
Krzyknąłem/Krzyknęłam/Krzyknął/Krzyknęła/Krzyknęło/Krzyknęliśmy/Krzyknęli
Wykrzyczałem/Wykrzyczałam/Wykrzyczał/Wykrzyczała/Wykrzyczało/Wykrzyczeliśmy/ Wykrzyknęli
Zakrzyknąłem/Zakrzyknęłam/Zakrzyknął/Zakrzyczwał/Zakrzyknęła/Zakrzyczwała/Zakrzyczwało/ Zakrzyknęło/Zakrzyknęliśmy/Zakrzyczeliśmy/Zakrzyknęli/Zakrzyczali

Wyekscerpowane metaoperatory można rozbudowywać według przyjętego sposobu, aczkolwiek wymagają one rozstrzygnięcia kwestii częstości ich użycia i kolejno przydatności

w procesie wyszukiwania reproduktów. Dla zaprezentowania skuteczności metody, polegającej na odnalezieniu odtwarzalnych jednostek wielowyrazowych języka mówionego wykorzystane zostaną te formuły wskaźnikowe, które wykazały się największą częstością występowania w zasobach NKJP: *mówi/powiedział/dodał*. Połączenie wskazanego metaoperatora z myślnikiem dialogowym w zapytaniu pozwala zawęzić wyniki wyszukiwania do konkordancji, zawierających opis narratorski wypowiedzi postaci i wydaje się wręcz konieczne przy tak dużych zasobach tekstowych. Kolejny etap analizy ilustrującej będzie stanowiła próba odnalezienia reproduktów potocznych przy wykorzystaniu wskazanych uprzednio metaoperatorów narracyjnych.

- MÓWI

Wpisanie zapytania w wyszukiwarkę PELCRA korpusu NKJP w postaci „- *mówi*” przyniosło najwięcej, bo aż 84 tysiące poświadczeń w korpusie zrównoważonym. Zastosowanie dodatkowego narzędzia korpusu, czyli funkcji *profil* umożliwiło sprawdzenie frekwencji występowania danego metaoperatora w różnych rejestrach języka. Otrzymany wykres słupkowy pokazuje, że analizowana formuła wskaźnikowa występuje najczęściej w tekstach publicystycznych i beletrystycznych. Pomimo wykorzystania w analizie zbiorów korpusowych i wygenerowania przez system listy konkordancji, zawierających wyszukiwany przez nas leksem, z całą pewnością nie możemy uznać zaprezentowany sposób dochodzenia do reproduktów metodą automatyczną. Otrzymany spis stanowi podstawę do dalszego opracowania i dokładnej obserwacji tych ciągów wielowyrazowych, w przypadku których nasza samoświadomość językowa podpowiada, by zostały one dodatkowo zweryfikowane i włączone do określonej zbiorowości reproduktów potocznych. Na potrzeby przeprowadzanej analizy wybrano 5 poświadczeń z różnych źródeł tekstowych, w których wielowyrazowe jednostki znajdują się w bezpośredniej bliskości zadanego metaoperatora.

Niech zrobi ciszej albo da coś po angielsku. - <u><i>A co ci za różnica</i></u> – mówi Lewandowski i patrzy w swoje kończące się piwo.
- Czy nie można by go jakoś zostawić w Danii? - spytał jeszcze raz Lone. - <u><i>Z jakiej racji?</i></u> - mówi oburzona - przecież wszystkie instytucje orzekły, że jest u nas nielegalnie.
- <u><i>Bujać to my, ale nie nas</i></u> - mówi i puszcza do mnie oko. - W porządku panie doktorze. Rozgryzłem pana. Może się pan ujawnić.
- <u><i>Porządek musi być</i></u> - mówi oddziałowy - bo inaczej będzie tu istny dom wariatów! Domyśliłem się, że osilek nie cieszy się sympatią nie tylko wśród pacjentów, ale także wśród personelu.

Przechodzi drugi.

- *Daj spokój* - **mówi**. - Tylu tam masz, to co ci zależy, wsadź jeszcze jednego.

Pierwsze wyrażenie *A co ci za różnica* odnotowuje *Wielki słownik języka polskiego* w wariantcie *Co za różnica* i przypisuje jednostce następującą definicję: *pot.* ‘używane w sytuacji, gdy mówiący chce powiedzieć, że nie widzi różnicy między wskazanymi osobami, rzeczami lub zjawiskami, wszystkie je ocenia jednakowo’.

Uzyskany reprodukt podkreśla dodatkowo osobę poprzez zastosowanie zaimka osobowego, przenosząc jednocześnie stosunek wypowiadającego na wywołanie pożądanej reakcji u rozmówcy. Bez względu na to, że wyodrębniona jednostka wielowyrazowa posiada pojedynczą liczbę wystąpień w korpusie, można ją uznać za modyfikację formy podstawowej (*Co za różnica*) i skonstruować dla niej następujący schemat: *A co <komu> za różnica*.

Z kolei wyrażenie z *jakiej racji* zostało zarejestrowane przez korpus NKJP w liczbie 149 poświadczeń i jest używane w sytuacji, gdy ‘nadawca wyraża swoje oburzenie, pytając, na jakiej podstawie coś się dzieje, gdyż uważa, że jest to bezpodstawne’ (możliwa synonimia do *jakim prawem*).

Reprodukt *porządek musi być* również jest rejestrowany przez korpus w liczbie 55 poświadczeń i zwyczajowo jest używany dla poparcia czyichś działań lub pozytywnego odbioru opisywanej sytuacji.

Przykład reproduktu *bujać to my, ale nie nas* przynosi zaledwie 5 poświadczeń, ale jest rejestrowany w *Wielkim słowniku języka polskiego* z eksplikacją: *pot.* ‘używane w sytuacji, gdy mówiący kwestionuje prawdomówność swojego rozmówcy’.

Zaprezentowana interpretacja nie budzi wątpliwości w kwestii wiarygodności definicji i istnieje duże prawdopodobieństwo, iż dana jednostka wielowyrazowa reprezentuje przykład reproduktu o postaci kształtowo zamkniętej (brak wariantywności). W przypadku ostatniego z wyodrębnionych reproduktów na pierwszy plan wysuwają się problemy z wieloznacznością danej jednostki wielowyrazowej, którego semantyka nie daje się zamknąć w ramach jednego potencjału treściowego. Istnienie reproduktu *daj spokój* poświadcza ponad tysiąc konkordancji w korpusie zrównoważonym, a różnorodne konteksty użycia danego ciągu wielowyrazowego pozwalają przypisać mu następujące znaczenia:

- Przepraszam, nie wiedziałem. - Daj spokój , nie ma się czym przejmować. Tak jest od śmierci ojca i zdążyłem już do tego przywyknąć.	1) <i>grzecznościowa formuła używana w odpowiedzi na czyjeś przeprosiny lub usprawiedliwianie się</i>
- Czy pan zwariował! - krzyczał Stec. – Tu, w interesie!? Co panu na mózg padło! Psiakrew! Tu awantury... gówniarz.... No co jeszcze! Spokój, bo... Trzymali go mocno, dopóki się nie uspokoił. Jawora przemawiał łagodnie: - No, Romciu, daj spokój , to nie ma sensu, gdzie to kto widział tak...	2) <i>uspokojenie rozmówcy</i>
- Ale ja... - Daj spokój , nie wysilaj się. Niemądrze zrobiłaś przychodząc, dom jest pod obserwacją. Nie chodzi o ciebie.	3) <i>mówiący daje do zrozumienia nadawcy, żeby nie kontynuował wypowiedzi</i>
- Tylko mi się tu dobrze sprawuj - powiedział, grożąc jej palcem. - Żadnych romansów ze studentami. - Daj spokój , zawsze się wygłupiasz przy pożegnaniu. Czy raz nie możesz być poważny? Przecież długo cię nie będzie - odrzekła rozżalona.	4) <i>wyrażenie niezadowolenia z informacji przekazanej przez nadawcę.</i>

Naturalnie, przedstawione definicje zapewne nie wyczerpują możliwości znaczeniowych wyekscerpowanego wyrażenia i zasadne wydaje się przeprowadzenie dokładniejszej analizy, wyniki której winny być uwzględnione w opisie leksykograficznym. Wspomniany już *Wielki słownik języka polskiego* z jednej strony uwzględnia daną jednostkę, a z drugiej ogranicza definicję tylko do następującego wyjaśnienia: *pot.* ‘sformułowanie wyrażające brak zgody mówiącego na coś, jego dezaprobatę lub zniecierpliwienie’.

- POWIEDZIAŁ

Analiza kolejnego metaoperatora narracyjnego w postaci zapytania „- *powiedział*” dostarczyła ponad 35 tysięcy poświadczeń, z których zdecydowana większość pochodzi z prozy bądź publicystyki. Poniżej przedstawiono przykłady tych konkordancji, których struktura sygnalizuje obecność potencjalnego reproduktu:

- <u>Bez przesady</u> - powiedział krawiec Kujawski. Pan był mały berbeć, kiedy prasowałem mundury pańskiego tatusia.
- <u>Nie o to chodzi</u> - powiedział Kleinfeld po chwili milczenia. - Ale dlaczego kazał pan im się podciągać?
- <u>Dosyć tego</u> - powiedział ostro. - Niech pan złazi i zapomnijmy o tym.
- <u>Teraz wszystko jest jasne</u> - powiedział Geralt. - Smok prawdopodobnie był w tych jaskiniach od stuleci, pogrążony w letargu. Słyszałem o takich wypadkach. I tam też musi być jego skarbiec.

- *Tak myślałem* - **powiedział** i otrząsnął z siebie śnieg jak pies wodę. - W ostatniej chwili zdążyłem. Jeszcze pół godziny i nie znalazłbym śladów.

Zasoby korpusowe przynoszą ponad tysiąc przykładów użycia dwuwyrazowej formacji *bez przesady*, jednak możliwość wtrącenia danego wyrażenia do wypowiedzi wymaga skrupulatnego sprawdzenia kontekstów użycia danej jednostki. Ograniczając się jednak do replik, których frekwencja w korpusie wynosi 224 poświadczenia, można ustalić konkretne znaczenie semantyczne: *pot.* ‘nadawca mówi, że to, co zostało powiedziane, to za dużo i nie stanowi to dokładnego odzwierciedlenia tego, jak jest’.

Z kolei sens wyrażenia *nie o to chodzi* zawiera się w zaprzeczeniu odnośnie do tego, co powiedział rozmówca bądź stanowi wprowadzenie do wyjaśnienia jakiegoś zagadnienia, które – zdaniem mówiącego – jest błędnie rozumiane przez rozmówcę. Przedstawiona jednostka wielowyrazowa jest odnotowywana przez korpus w niespełna tysiącu poświadczeniach, jednak próby odnalezienia wyrażenia i jego semantyki w spisie leksykograficznym WSJP nie przyniosły rezultatu.

Z kolei sformułowanie *dosyć tego* służy dla wyrażenia zdenerwowania rozmówcy i podkreślenia braku akceptacji w stosunku do tego, co powiedział rozmówca. Opis uzupełniający definicję, znajdujący się pod hasłem *dosyć* w WSJP brzmi następująco: ‘mówiący uważa, że to, o czym mowa trwało zbyt długo i chce, żeby się skończyło’. Analizowany wielowyrazowiec posiada w korpusie 215 poświadczeń i w stosunku do wariantu *dość tego* używany jest znacznie rzadziej. Nie jest uwzględniany w źródłach systemowych¹⁰.

Kolejny reprodukt potoczny *tak myślałem* potwierdza wcześniejsze przypuszczenia mówiącego w stosunku do komunikatu rozmówcy, a obserwacja 270 kontekstów wykazała, iż reprodukt cechuje się zróżnicowanym nacechowaniem emocjonalnym, np. rezygnacją, rozczarowaniem, ekscytacją itp.

Jednostka *teraz wszystko jasne* jest rejestrowana w zaledwie 7 kontekstach użycia, co najpewniej wynika z preferencji użytkowników do odtwarzania w komunikacji formy wyjściowej reproduktu, tj. *wszystko jasne*:

1)	<i>Było tam wydrukowane grubą czcionką: PRZYGODA ZE ŚMIERCIĄ... Tak, teraz wszystko jasne!</i>
2)	<i>- Jestem drużbą Tadeusza. Aha, tak wygląda mój weselny partner. Teraz wszystko jasne. Znam go z opowiadań przyszłego szwagra, podobno sympatyczny chłopak, teraz mogę poznać naprawdę.</i>

¹⁰ Przez źródła systemowe należy rozumieć bezkontekstowe zbiory jednostek języka (termin używany za Wojciechem Chlebą).

3)	<i>Twarz Hansa się rozpogodziła. - Aaa... teraz wszystko jasne: to był Niemiec nadwołżański! Proboszcz wesolo przeczy głowę.</i>
4)	<i>- Tak – szepnęła. – Niewątpliwie tego zażądają. - No pewnie. Teraz wszystko jasne. Nikt nikomu nie daje, ot tak sobie, dwudziestu kawalków miesiąc w miesiąc.</i>
5)	<i>- Na urlopie oświadczył się i został przyjęty - wyjaśnił podporucznik Kawiarz. - Teraz wszystko jasne - uśmiechnął się major Milewski.</i>
6)	<i>- Niech będzie na cztery – warknęła, patrząc na mnie ze złością. - No, teraz wszystko jasne - skinąłem głową i pochyliłem się nad kartką.</i>
7)	<i>- Rozumiem! Mój organizm tego nie wymaga. W moim świecie kobieta sama karmi swoje dzieci, więc zarodek nie potrzebuje sygnałów od ojca [...]. - Ach, tak! Teraz wszystko jasne. Kobiety, same karmiące dzieci po ich urodzeniu!</i>

W przypadku wyodrębnionej jednostki wielowyrazowej podkreślona zostaje chwila zrozumienia sytuacji, opisywanej przez nadawcę wypowiedzi, co pozwala na skonstruowanie konkretnego schematu: <teraz> wszystko jasne.

- DODAŁ

Ostatni z zaproponowanych metaoperatorów narracyjnych występuje w liczbie ponad pięciu tysięcy poświadczeń w korpusie zrównoważonym NKJP, a jego wykorzystanie w procesie poszukiwania jednostek reprodukowalnych utrudnia niestety jego lokalizacja w środku wypowiedzi. Odszukanie więc wyrażenia, stanowiącego reakcję na komunikat rozmówcy w postaci komentarza uzupełniającego wymaga ściśle szczegółowej analizy uzyskanych konkordancji, co zdecydowanie obniża przydatność danego metaoperatora przy wyodrębnianiu potencjalnych reproduktów. Wyselekcjonowane konteksty użycia przedstawiają się następująco:

- <u>I jeszcze jedno</u> - dodał Zani, ściskając prawicę Wyższego. - Dopilnujesz, żeby trzymano ich razem, i nie pozwolisz na żadne przesłuchania.
<u>Ale numer</u> - dodał ciszej i odetchnął z ulgą. Wcale nie miałem ochoty na żarty.
Nawiasem mówiąc, jabłka i brzoskwinie zalicza się do symboliki narządu płciowego żeńskiego. <u>Liczę na to</u> - dodał doktor Hogarth - iż daruje mi pan teraz przesadne określenie.
- <u>Właśnie tak!</u> - dodał Darrell. - I jak pan widzi, dobrze zrobiłem. Po co wam nasza dokumentacja? Wszystko, co powinniście wiedzieć, wynika z porozumień pomiędzy naszymi rządami.
- <u>Proszę sobie wyobrazić</u> - dodał jeszcze - że dwa dni temu dostałem podobny list z Indii: "Będziesz płacił za winy twego stryja. Nie ma przebaczenia!". Oto ten list.

Sformułowanie *I jeszcze jedno* występuje jako dopełnienie nadawcy do przekazywanej przez rozmówcę wiadomości i w jego ocenie stanowi ważny aspekt uzupełniający treści

komunikatu. W zaprezentowanym kontekście reprodut nie tylko podtrzymuje dalszy ciąg rozmowy, pełniąc zarazem funkcję fatyczną wypowiedzi, ale jest swego rodzaju nakazem wykonania żadanego polecenia. W zasobach korpusowych posiada aż 957 poświadczeń użycia.

Za pomocą reprodutu *Ale numer* mówiący wyraża swoje zdziwienie w stosunku do opisywanej przez rozmówcę sytuacji i stanowi on komentarz podsumowujący. Pomimo swojej wysokiej frekwencji w języku potocznym nie jest on uwzględniony w *WSJP*, a korpus zrównoważony rejestruje 122 poświadczenia.

Nadzieje co do pozytywnego rozwoju sytuacji wyraża wielowyrazowiec *liczę na to*, który jest odnotowany 361 razy. Jego struktura uwzględnia zarazem przypisaną reprodutowi wariantywność, którą można przedstawić następująco: *liczę na <kogo/co>*.

Wyrażenie *właśnie tak* w akcie komunikacji należy traktować jako wykładnik zgody, a dokładną jego definicję znaczenia przytacza leksykon *WSJP*: ‘mówiący zgadza się z tym, co zostało powiedziane, uważając, że to, a nie coś innego, powinno być powiedziane o tym, o czym mowa’. *NKJP* rejestruje niespełna 1500 konkordancji danego reprodutu.

Przykład ostatni *proszę sobie wyobrazić* stanowi zapowiedź do przekazywanej przez mówiącego wiadomości, która może wydać się adresatowi komunikatu dziwna czy zaskakująca. Przytoczony reprodut posiada 493 poświadczenia i przeważnie jest używany w stosunku do osoby, z którą nie łączą mówiącego relacje bliższe w odróżnieniu od sformułowania *wyobraż sobie*, które liczy 1302 konkordancje.

Czeskie metaoperatory narracyjne

Analogiczną metodę możemy zastosować dla opracowania indeksu czeskich metaoperatorów narracyjnych i określeniu stopnia ich przydatności w procesie poszukiwawczym reprodutów języka mówionego. W odróżnieniu od przedstawionego uprzednio schematu sygnalizowania komunikatu w polskiej literaturze, czeski model winien zawierać znak prawego podwójnego cudzysłowu " (zamiast myślnika dialogowego), który w czeskiej literaturze zastępuje pauzę dialogową. Po wprowadzeniu wskazanego symbolu do zbiorów korpusu czeskiej wersji *InterCorp v16ud* otrzymujemy ponad milion konkordancji, każdorazowo wymagających analizy struktury i treści znajdujących się w bliskim sąsiedztwie symbolu podanego w zapytaniu. W celach egzemplifikacyjnych wykorzystano ponad tysiąc konkordancji, z których izolowano metaoperatory narracyjne wskazane w poniższej tabeli:

VÝCHOZÍ FORMA	NARATIVNÍ METAOPERÁTOR
Breptat	Breptal/Vybreptali
Dodat	Dodal jsem/Dodal/Dodala
Křičet	Křičím/Vykřičela
Mluvit	Mluvil/Mluvila
Myslet	Myslím/Myslela/Mysleli
Odpovědět	Odpověděl jsem/Odpověděl/Odpověděla
Opakovat	Opakuju/Opakoval/Opakovala
Ozvat se	Ozval se/Ozvala se
Prohlásit	Prohlásil/Prohlásila jsem
Ptát se	Ptám se/Zeptal jsem se/Zeptala se
Říct	Řekl jsem/Řekl/Řekla
Ujistit se	Ujistil jsem se/Ujistila se
Vzdychnout	Vzdychl jsem/Vzdychl/Vzdychla
Zajímat se	Zajímal se/Zajímala se
Zajistit	Zajistil jsem/Zajistil/Zajistila
Zavolat	Zavolal jsem/Zavolal/Zavolala

Spośród zaprezentowanego wykazu wyekscerpowanych metaoperatorów narracyjnych, najwyższą frekwencję użycia w źródłach wspomnianego korpusu InterCorp można przypisać jednostkom: *řekl* i *odpověděl*. Połączenie wskazanego metaoperatora ze znakiem (") w zapytaniu pozwala zawęzić wyniki wyszukiwania do konkordancji zawierających narracyjny opis wypowiedzi nadawcy, a jego rezultaty przedstawiają się następująco:

" ŘEKL

Wpisanie zapytania do wyszukiwarki korpusowej KonText korpusu InterCorpv16 ud w postaci "*řekl*" dostarczyło największej liczby poświadczeń w czeskich źródłach tekstowych, wynoszącej 47 tysięcy konkordancji. Zastosowanie fakultatywnego narzędzia korpusowego, czyli funkcji *typ tekstu*, umożliwiło sprawdzenie częstości występowania danego metaoperatora w różnych dokumentach tekstowych, wśród których najwięcej poświadczeń zawierają się w tekstach beletrystycznych (czes. *jádro*). Poniżej przedstawiono cztery fragmenty tekstowe, w których zaobserwowano występowanie wskazanego metaoperatora oraz potocznej jednostki wielowyrazowej.

<i>Už nikdy se nebude dívat z Chrisovy terasy na moře, nikdy už nepoběží po břehu. Už nikdy ...Všechno je pryč! "<u>Co se má stát, stane se,</u> " řekl Chris. Měl pravdu.</i>
<i>Nikdy jsem sice nepracovala přímo v baru, ale jednou jsem dělala ve školní jídelně, takže mýt nádobí umím. A roky jsem pracovala na pokladně v knihkupectví.“ „<u>Já o tom nepochybuju,</u> “ řekl Tom.</i>
<i>Pokusila se o smích, ale vydala ze sebe jen napůl potlačené vzlyknutí. Sevřel ji pevněji. „<u>To se vyřeší,</u> “ řekl, nejspíš proto, že nepřišel na nic jiného, co by mohl říct.</i>
<i>Já znal kněze, kteří se převlékali do občanských šatů a chodili se dívat, jak tanečnice mrskají nohama. "<u>To snad ne!</u> " řekl farář. "Ano, já takové znal!</i>

Pierwsza z odnalezionych fraz *Co se má stát, stane se* jest odnotowana 214 razy w korpusie, których analiza semantyczna pozwala na przypisanie jednostce następującej definicji: „Mówiący, który nie zna przyszłych konsekwencji informacji przekazanych przez rozmówcę, wyraża brak wpływu na przebieg omawianej sytuacji oraz akceptację jej następstw.”

Wyrażenie *Já o tom nepochybuju* stanowi komentarz dorematyczny, który można interpretować jako przekonanie mówiącego o prawdziwości swojego sądu i pewności w stosunku do przekazanej mu informacji. Ów reproduct odznacza się jednak niską frekwencją użycia w korpusie (8 poświadczeń), co jest najprawdopodobniej spowodowane preferencją użytkowników do stosowania formy *Nepochybuju o tom*. Trzecia z przytoczonych fraz *To se vyřeší* stanowi komentarz, który ma na celu uspokoić nadawcę wypowiedzi i zachęcić go do ostudzenia emocji. Dany reproduct został zarejestrowany w korpusie w 64 przykładach, najczęściej służy do sygnalizacji zakończenia podejmowanego wątku wypowiedzi i tworzy podstawę dla rozpowszechnionego wariantu *To se vyřeší samo*. Ostatni wieloznaczny wielowyrzowiec *To snad ne!*, który został odnotowany w największej liczbie 4637 konkordancji, został w przedstawionym kontekście użyty w celu wyrażenia zaskoczenia mówiącego w reakcji na otrzymaną informację, oraz niedowierzania w słowa mówiącego.

" ODPOVĚĎĚL

Kolejny z wyekscerpowanych metaoperatorów narracyjnych jest odnotowywany w korpusie w liczbie 8632 poświadczeń, z których aż 8560 wystąpień zawiera się w źródłach tekstów beletrystycznych. Selekcja wybranych konkordancji pozwoliła wyodrębnić poniższe poświadczenia tekstowe, w których zastosowano wskazany metaoperator i wielowyrzową jednostkę potoczną:

„řekl rychle Hlavoun, „ty půjdeš se mnou a otevřeš kotec. Vy ostatní tu dávejte pozor. Jestli se tu objeví další kočka, budete si s ní muset poradit sami. “ „<u>Beze všeho</u>, “ odpověděl Pampeliška.
Ale nemám chuť poslat s ním dceru přes moře, a dostat zetě z Grimelundaru se mi taky nechce. “ „ <u>Není se co divit</u>, “ odpověděl Veterlide a dál už o té věci nemluvili.
„Možná. “ „ Takže pak budeš mít další dítě. “ Přikývla. „A ty dostaneš bratříčka. Zamlouvá se ti to? “ „<u>Je mi to jedno</u>, “ odpověděl. „ Tom už mi tě tak jako tak vzal.
Líbily se mi její oči. „Odkud to všechno víš? “ zeptal jsem se Samuela v domnění, že se třeba živil jako učitel dějepisu. „ <u>Těžko říct</u>, “ odpověděl Samuel a usmál se. „Myslím, že to mám z nějaký počítačový hry.

Wyrażenie *beze všeho*, którego różne znaczenia aktualizują się w zależności od konkretnej sytuacji komunikacyjnej, zostało zarejestrowane w korpusie w 34 przykładach. Znaczenie można sprowadzić do następującej definicji: ‘omawiana kwestia jest postrzegana jako prosta, niewymagająca szczególnego wysiłku czy umiejętności, a jej realizacja jest w pełni możliwa do osiągnięcia w łatwy sposób’. W pozostałych kontekstach służy również do wyrażenia zgody, akceptacji czy chęci spełnienia prośby rozmówcy, np.

Můžeš mi s tím pomoci? - Beze všeho;

Můžeme si od tebe půjčit tu knihu? - Beze všeho.

Stwierdzenie *není se co divit* występuje w InterCorp w 97 kontekstach i jest stosowane do wyrażenia, że opisywana sytuacja lub zdarzenie są przewidywalne lub zrozumiałe w kontekście zaistniałych i znanych mówiącemu okoliczności. Służy do podkreślenia, że przebieg i rezultat określonego zdarzenia jest naturalnym wynikiem pewnych warunków bądź zachowań i nie powinno powodować zaskoczenia. Warto nadmienić, iż wyższą odnotowaną frekwencją użycia charakteryzuje się wariant danego wielowyrzowca w postaci *není se čemu divit*.

Jeden z najpowszechniejszych reproductów reakcyjnych, wyrażenie *je mi to jedno*, rejestruje się w korpusie w 3379 konkordancjach i służy do zasygnalizowania rozmówcy, że przekazywana informacja nie budzi u mówiącego zdziwienia i jest odbierana obojętnie. Synonimiczną jednostką, charakteryzującą się wyższym stopniem potoczności niż wskazany reproduct, jest pokrewne wyrażenie *je mi to fuk*.

Ostatni wielowyrzowiec *těžko říct* (1285 wystąpień) używany jest w sytuacjach, gdy odpowiedź na pytanie jest niejednoznaczna, skomplikowana bądź niepewna. Wyraża również brak wiedzy na dany temat i niekompletność potrzebnych informacji dla udzielenia trafnej odpowiedzi, np.

*Myslíš, že zítra bude pršet? - Těžko říct, předpověď je nejistá;
Jak tohle všechno dopadne? - Těžko říct.*

Rosyjskie metaoperatory narracyjne

Przedstawiona powyżej metoda wyszukiwania i ekscepcji metaoperatorów narracyjnych komplikuje się w przypadku prób jej zastosowania w NKJR. W rosyjskiej literaturze beletrystycznej myślnik dialogowy jest zastąpiony znakiem pauzy — (ros. *Тупе*), którego wpisanie w wyszukiwarkę korpusu NKJR nie przynosi żadnych rezultatów z powodu braku możliwości zadania zapytania w formie samodzielnego symbolu. Można więc założyć, iż najskuteczniejszą analogiczną metodą byłoby tłumaczenie uzyskanych uprzednio metaoperatorów narracyjnych polskich na język rosyjski, wpisanie ich w formularz zapytania wraz z odpowiednim symbolem i ustalenie stopnia frekwencji użycia dla każdego z nich. Jednakże już pierwsza z przeprowadzonych analiz, która zakładała wyszukanie reproduktów potocznych przy metaoperatorze —*сказал*, pozwoliła zaobserwować fakt pomijania przez korpus symbolu pauzy w większości poświadczeń tekstowych, co uniemożliwia określenie częstości wystąpienia danego metaoperatora w dostępnych zasobach i czyni tę metodę wysoce czasochłonną. Rozwiązaniem tego problemu jest zastosowanie dokładniejszej składni zapytania, zakładającej jednoczesne wpisanie w polu *Словоформа* leksemu *сказал* z uwzględnieniem dodatkowej funkcji (*дополнительные признаки*) w wyborze *Слово после* → *тупе* (adash). Po zastosowaniu danej procedury korpus ekscerpuje wyłącznie te poświadczenia, które zawierają wskazaną przez użytkownika formę —*сказал*:

Слово 1



Лемма



Словоформа



сказал

Грамм. признаки

выбрать



Семантика

выбрать



☒ 1-е знач. ☒ др. знач.

Синтаксические
свойства слова

выбрать



Доп. признаки

выбрать



adash

добавить свойство



☒ Слово после

- ☐ любого знака препинания
- ☐ точки
- ☐ запятой
- ☐ двоеточия
- ☐ точки с запятой
- ☒ тире
- ☐ дефиса
- ☐ восклицательного знака
- ☐ вопросительного знака

Запрос

Вернуться к поиску

• 10 880 текстов
• 163 160 примеров

предпочтительные разборы, сосе...



Добавить
в сравнение



Сохранить
запрос

сказал, adash

Конкорданс KWIC График Статистика Частотность 1-граммы 2-граммы 3-граммы ▾



Скачать ▾



1. Форум: 17 мгновений весны (2005-2010) 🗨

Рад видеть вас, чертей, — **сказал** Мюллер, — снова затеваете какое-нибудь очередное коварство? 🗨 ↔

2. В. Э. Карпов, Т. В. Мещерякова. Об автоматизации нетворческих литературных процессов // «Информационные технологии», 2004 🗨

— Спасибо за фоторобот, — **сказал** Шерлок Холмс секретарю. 🗨 ↔

— А вот что есть у меня! — **сказал** Шерлок Холмс и предъявил профессору Мориарти улику. 🗨 ↔

— А вот что есть у меня! — **сказал** Шерлок Холмс и предъявил профессору Мориарти улику. — Ладно, делать нечего, — **сказал** профессор Мориарти. — Держи. 🗨 ↔

— Благодарить не буду, — **сказал** сурово Шерлок Холмс профессору Мориарти. 🗨 ↔

3. Ольга Зуева. Скажи, что я тебе нужна... // «Даша», 2004 🗨

Плохо мне с ней, — **сказал** он и, подумав, добавил: — Во всех областях. 🗨 ↔

4. Павел Куприянов. Тем же курсом // «Computerworld», 2004 🗨

" Нам не хватает людей, мы даже начинаем переманивать их друг у друга ", — **сказал** Щербаков. 🗨 ↔

Za etap następny należy przyjąć ustanowienie rosyjskiego ekwiwalentu dla wybranych polskich metaoperatorów, który umożliwi określenie ich częstości użycia (poprzez zastosowanie symbolu pauzy „—” z metaoperatorem narracyjnym). Poniżej zaprezentowano tłumaczenie wybranych metaoperatorów narracyjnych, wśród których wyodrębniono sześć jednostek, charakteryzujących się największą liczbą poświadczeń w zbiorach NKJR:

ИСХОДНАЯ ФОРМА	ПОВЕСТВОВАТЕЛЬНЫЙ МЕТАОПЕРАТОР
Добавить (dodać)	Добавил/Добавила
Говорить (mówić)	Говорю/Говорит
Сказать (powiedzieć)	Сказал/Сказала
Повторить (powtórzyć)	Повторяю/Повторил/Повторила
Убедиться (upewnić się)	Убедился/убедились
Вздохнуть (westchnąć)	Вздохнул/Вздохнула
Заинтересоваться (zainteresować się)	Заинтересовался/Заинтересовалась
Уверить (zapewnić)	Уверил/Уверила

Повествовательный метаоператор	Число вхождений
—Сказал / —Сказала	163 160 / 48 230
—Говорю / —Говорит	9601 / 39 044
—Думал / —Подумал	8306 / 14 231

Powyższe dane liczbowe potwierdzają zaobserwowane podczas poprzednich analiz wnioski, które identyfikują metaoperator *—powiedział/" řekl/—сказал* jako jeden z najbardziej frekwencyjnych wskaźników tekstowych, umożliwiających odnalezienie reproduktów potocznych. W przypadku niektórych ze wskazanych rosyjskich metaoperatorów nie należy zapominać o możliwości ich występowania i pokrycia homonimicznego zarówno w trzeciej, jak i w pierwszej osobie liczby pojedynczej, co narzuca konieczność dodatkowej analizy wyekscerpowanych przez korpus poświadczeń (np. —*Что ты думаешь?*—сказал я [1 os. l. poj]; —*В том случае бывает по-разному* —сказал Михаил [3 os. l. poj]). W celu podjęcia próby wyzyskania rosyjskich wielowyrazowych jednostek języka mówionego zostaną wykorzystane wybrane dwa metaoperatory, czyli —*Сказал* i —*Говорит*.

—СКАЗАЛ

Wyekscerpowane poświadczenia w liczbie 163 160 kontekstów użycia zostały zarejestrowane głównie w zbiorach literatury pięknej w formie opowiadań o tematyce realiów życia codziennego (uzyskanie wskazanych danych umożliwiają dostępne funkcje korpusu: *сфера функционирования, тематика, тип текста*). Zastosowanie fakultatywnego narzędzia „2-граммы” pozwoliło zaobserwować, iż najczęstszym elementem występującym w bezpośredniej bliskości zadanego metaoperatora, jest zaimek osoby dla trzeciej osoby liczby pojedynczej rodzaju męskiego, tj. *сказал он* (37 594 przykładów użycia). Owa procedura minimalizuje więc czasochłonność każdorazowego określania konkretnej osoby gramatycznej przy metaoperatorze —сказал. Spośród udostępnionych 10 880 typów tekstów wybrano poniższe przykłady tekstowe, w których zarejestrowano jednostki wielowyrazowe o wysokim prawdopodobieństwie regularnej reprodukowalności w języku Rosjan:

*Но снова дымок беспокойства легко растаял в обширном пространстве его усталости и тоски. Варя отворила дверь. Малиновский, виновато поглядывая, стаскивал ботинки. — **Без разговоров**, — сказал он, — ком цу мир...*

*Он о чем-то задумался... Потом потянулся к мужику, у которого жиклер засорился, а мотор работает. — **Дело в том**, — сказал он, — что это не жиклер засорился! Понял?! — А что же?*

*— **Легко вам говорить**, — сказал я, — у вас во-о-он хозяйство какое: и огурчики, и шуба, и трактор...*

*— А я играю, — живо возразил Красноперов, — за честь института. — **Так я и думал**, — сказал незнакомец, — всех благ!*

W przypadku pierwszego odnotowanego wyrażenia *без разговоров* na pierwszy plan wysuwają się problemy, związane z możliwością zastosowania owej frazy w znaczeniu dosłownym, tj. *bez rozmów* (np. *Путь продолжали, но без разговоров, думая каждый о своем*), co uniemożliwia jednoznaczne wskazanie frekwencji reproduktu w znaczeniu reakcyjnym. Definicję wskazanego reproduktu uwzględnia *Фразеологический словарь русского литературного языка*, który odnotowuje jednostkę wraz z najczęstszą realizacją fakultatywną → БЕЗ РАЗГОВОРОВ: БЕЗ <ВСЯКОГО> РАЗГОВОРА (РАЗГОВОРУ). Разг. Экспрес. 1. *Без возражений, не высказывая недовольства* (Фёдоров 2008: 559).

Za wysoce prawdopodobny ekwiwalent polski można by było uznać frazę *Bez dyskusji*, semantykę której podaje *WSJP*: ‘używane dla podkreślenia, że stanowisko mówiącego w danej kwestii nie ulegnie zmianie i dlatego należy się bez zbędnych sprzeciwów podporządkować’.

Wielowyrazowiec *Дело в том* (pełniący funkcję fatyczną, wprowadzającą do komunikatu) jest odnotowywany w 13 240 poświadczeniach tekstowych, na podstawie których można przypisać mu następujące znaczenie: ‘najważniejszy aspekt danej sytuacji to ten, który zostanie poruszony w dalszej części komunikatu mówiącego’.

Należy zauważyć, że dane wyrażenie jest tłumaczone na język polski dwojako: *rzecz w tym/chodzi o to, że...*, co czyni zasadnym uwzględnienie dodatkowej możliwości zastosowania danego reproduktu w następującej sytuacji komunikacyjnej: ‘to, co zostało wspomniane na dany temat, można doprecyzować w następujący sposób...’.

Trzeci przykład reproduktu z wyekscerpowanego poświadczenia tekstowego zawiera realizację miejsca walencyjnego, tj. *Легко <кому?> говорить*, a jego forma kanoniczna jest zarejestrowana w 158 kontekstach użycia. Analizowane wyrażenie charakteryzuje się zdecydowanie niższą frekwencją użycia niż synonimiczny zwrot *легко сказать* (644 poświadczenia), którego semantykę można sprowadzić do następujących interpretacji: ‘употребляется как ответ на предложение, требование предпринять, сделать что-л. очень трудное или невозможное’/ *Возражение по поводу чего-либо, что представляется сложным, трудным*.

Ostatnia z przedstawionych jednostek wielowyrazowych *так я и думал* jest stosowana w celu wyrażenia wcześniejszych domysłów mówiącego w kontekście do zaistniałej sytuacji komunikacyjnej (107 kontekstów użycia). W charakterze frazy reakcyjnej służy podkreśleniu zgody z tym, co zostało wypowiedziane przez rozmówcę. Należy zaznaczyć, iż wyodrębniony z zasobów korpusowych reprodukt stanowi wariant pokrewnej jednostki wyjściowej, charakteryzującej się wyższym stopniem rozpowszechnienia wśród rodzimych użytkowników języka rosyjskiego, tj. *я так и думал*.

—ГОВОРИТ

Metaoperator narracyjny *—говорит* został zarejestrowany w liczbie 39 044 poświadczeń tekstowych, z których największy procent stanowią artykuły publicystyczne, poświęcone polityce i życiu codziennemu. Ponowne zastosowanie funkcji „2-граммы” przynosi analogiczne do poprzedniej analizy obserwacje, które wskazują 3. os. l. p. r. m. jako element, występujący najczęściej w bezpośredniej bliskości zadanego metaoperatora (*он*

говорит). Wyniki ręcznej ekscerpacji fragmentów kontekstowych, zawierających jednostkę wielowyrazową i omawiany wskaźnik tekstowy, prezentują się następująco:

— Значит, я должен плюнуть на аспирантуру ради ваших интересов? — Нам учителя нужны. — <u>А мне какое дело?</u> — говорит он. Поле боя покрыто пороховым дымом.
Уже половина шестого, но жара всё такая же удушающая, и отражающий солнце подоконник, наверное, раскалённый, как доменная печь, нарисованная в школьном учебнике литературы. — <u>Мне пора</u> , — говорит мама и снимает фартук.
Линия, по-моему, должна быть только одна ". — " Ты так считаешь? " — спрашивает Рита. " <u>А как же иначе!</u> " — говорит Лариса, изумляясь и возмущаясь одновременно тем, что могут быть какие-либо сомнения.
— <u>Вот и все</u> , — говорит Ада, распрямляет спину и устало проводит по лицу зеленым рукавом, стирая испарину. — Гелле! Пробегись-ка по лестнице, попробуй, что получилось. Близится ночь.

Pierwszy z wyodrębnionych reproduktów <A>мне какое дело został odnotowany w zasobach korpusowych w liczbie 113 poświadczeń, a zawarte w słownikach znaczenia danego wyrażenia można zawrzeć w następującej definicji: ‘Данная фраза чаще всего выражает равнодушие, безразличие или пренебрежение к чему-то, а её точное толкование зависит от контекста использования выражения’.

Zdecydowanie częściej obserwuje się użycie jednego z wariantów danego wielowyrazowca, tj. *какое мне дело* (411 poświadczeń), a słownik synonimów autorstwa W. N. Triszina podaje synonimiczną reproduktowi jednostkę *мне какая печаль*, którą rejestrują wyłącznie pojedyncze przykłady użycia w zasobach rosyjskiego korpusu.

Konstrukcję zwrotu *мне пора*, który został zastosowany w ponad 1000 poświadczeniach tekstowych, można sprowadzić do modelu, uwzględniającego wymiennność miejsca walencyjnego w postaci <кому> *пора*. Interpretacja tego zwrotu w sposób wyczerpujący jest przedstawiona w pracy *Толковый словарь русского языка Дмитриева*: ‘Когда говорят, что кому-то пора что-то сделать, имеют в виду, что настало время, срок, когда необходимо это сделать’. Należy zauważyć, że samodzielne użycie leksemu *пора* + *инфинитив* (np. *пора спать*) służy każdorazowo wyrażeniu przez mówiącego konieczności zakończenia trwającej czynności i potrzebie rozpoczęcia innej. Z kolei wielowyrazowa jednostka reakcyjna <A> *как же иначе*, która liczy ponad 700 przykładów tekstowych w zasobach korpusowych, stanowi rodzaj pytania retorycznego, które podkreśla naturalność

określonego działania w konkretnej sytuacji i jedyną słuszną możliwość jego zastosowania. Jednocześnie służy wyrażeniu pewności mówiącego co do słuszności wykonania zaproponowanej czynności bądź brak innych alternatyw dla rozwiązania wskazanego problemu. Równie często jest odnotowywana skrócona forma danego wielowyrazowca, która pomija partykułę *же* i przyjmuje formę *а как иначе*. W przypadku ostatniego zaprezentowanego wyrażenia *вот и всё* nie należy sugerować się odnotowaną w korpusie liczbą 5660 poświadczeń z uwagi na to, że rosyjskojęzyczne zasoby tekstowe nie uwzględniają znaku diakrytycznego litery *ѐ*, co doprowadza do występowania homografów o innej wymowie, tj. *все/все* (w domyśle *всѐ*). Automatyczne wyekscerpowanie liczby kontekstów użycia re produktu *вот и всё* uniemożliwia występowanie następującego rodzaju przykładów tekstowych: *Нет, какое там божество: в одной стране придумали, в другой перенимают — вот и все причины*. Znaczenie danej jednostki wielowyrazowej jest odnotowywane przez różne opracowania jednojęzyczne z zakresu leksykografii rosyjskiej, np. *част. Употребляется при подведении итога сказанному, при указании на исчерпывающий характер сказанного* (Современный толковый словарь русского языка); *разг. неизм. На этом конец (после перечисления указывает на то, что оно закончено)* (Учебный фразеологический словарь). Do najbardziej rozpowszechnionych polskich tłumaczeń danego wielowyrazowca można zakwalifikować następujące jednostki: *to wszystko; to już koniec; to tyle; to by było na tyle* itp.

Представiona metoda wyodrębniania metaoperatorów narracyjnych i ich dalszego wykorzystania w procesie poszukiwawczym jednostek reprodukowalnych języka mówionego niewątpliwie wymaga dokładniejszego opracowania, w tym ustalenia kryteriów trafniejszej selekcji formuł wskaźnikowych. Niemniej jednak zaprezentowane wprowadzenie do omawianej problematyki sygnalizuje konieczność stworzenia nowego typu metaoperatorów i wyekscerpowania w szczególności tych jednostek, wykazujących się dużą przydatnością przy nieautomatycznych metodach dochodzenia do re produktów. Wygenerowane przez korpus listy konkordancji stwarzają ogromne możliwości dla typologizacji zawartych w nich metaoperatorów, których dalsze zastosowanie w zapytaniu doprowadza do odnalezienia jednostek wielowyrazowych, wymagających ustalenia postaci kształtowej i znaczenia semantycznego. Zaprezentowany sposób pozwolił wyodrębnić z poświadczeń 31 jednostek, podejrzanych o regularną odtwarzalność w mowie codziennej bądź już zawierających się w opracowaniach leksykograficznych. W niektórych przypadkach zaobserwowano niską liczbę kontekstów użycia lub nieobecność wskazanych re produktów w innych źródłach systemowych, co jednak nie powinno być powodem przedwczesnego odrzucenia wyzyskanego materiału dla

dalszej analizy. W przytoczonej metodzie uwidacznia się także aktualna problematyka wariantowości reproduktów, której szczegółowe opracowanie pozwoliłoby uniknąć dylematów, dotyczących poprawności i wiarygodności wyodrębnionej jednostki wielowyrazowej.

6.4 Metoda synonimiczna

Pierwszą z przedstawionych metod wyszukiwania i identyfikacji nowych jednostek wielowyrazowych lub ich wariantów, opartą na uprzednio zgromadzonej bazie reproduktów potocznych (fatyzmów i emotywów), stanowi tzw. **metoda synonimiczna**. We wszystkich analizowanych przez nas językach interpretacja terminu synonim wydaje się zbieżna i spójna, o czym mogą świadczyć poniższe wnioski w zakresie ich definicji:

„Synonimia to relacja implikacji dwustronnej między znaczeniami wyrażen. Wyrażenia A i B są synonimami, jeżeli znaczenie A implikuje znaczenie B i znaczenie B implikuje znaczenie A.” (Grochowski 1982: 73).

«Синонимы — это группа слов, за счет которой обогащается лексический состав русского языка. Это слова совпадающие или близкие по значению, но отличающиеся по написанию и звучанию. В тексте и речи синонимы выполняют одну и ту же грамматическую и синтаксическую функцию. Особенность синонимов заключается и в том, что даже имея одинаковое значение, они отличаются по экспрессии, употребительности, а также закреплённости за тем или иным стилем» (Гаврилов 2014: 3).

„Z hlediska stylistiky jsou [...] synonyma takové paralelní pojmenovací jednotky, které se na základě společného významového jádra v jistém kontextu buď zaměňují, nebo sbližují; z hlediska lexikologie to jsou lexikální jednotky se stejnou nebo odstíněnou významovou složkou pojmovou nebo i expresivní a se stejnou nebo částečně odlišnou oblastí kontextového užití.” (Poldauf 1962: 306).

Na podstawie powyższych informacji można zatem stwierdzić, że synonimem jest jednostka, umożliwiająca zastąpienie danego wyrazu bez znaczącej zmiany treści wypowiedzi, choć może się różnić pod względem nacechowania stylistycznego czy frekwencji

występowania w języku. Należy jednak zauważyć, że w przeważającej większości dostępnych opracowań leksykograficznych zamieszczane są wyłącznie przykłady synonimów jednowyrazowych, przy czym proponowane zamienniki również ograniczają się do form jednowyrazowych np. *wróżba – przepowiednia, prorocstwo; брак - бракосочетание, венчание; rodina - příbuzenstvo, familie*. Wyraźna koncentracja twórców słownika na najbardziej neutralnych i bezpośrednich ekwiwalentach danego słowa, eliminuje możliwość ujęcia jednostek bardziej złożonych (wielowyzrazowych) i ich wieloelementowych odpowiedników. W danym aspekcie kluczowym z punktu widzenia badań nad fatyzmami i emotywami jest pytanie: Czy dana jednostka wielowyzrazowa zostałaby użyta w identycznej sytuacji komunikacyjnej i niosłaby podobny potencjał treściowy? Załóżmy, że zależy nam na wyrażeniu obojętności w sposób bardziej zróżnicowany niż za pomocą podstawowego i najczęściej używanego wyrażenia *nie obchodzi mnie to/мне всё равно/je mi to jedno*. W języku mówionym mamy do dyspozycji szereg synonimicznych jednostek reakcyjnych, które służą wyrażaniu podobnego stosunku, lecz niewątpliwie odznaczają się różnym stopniem emocjonalnego nacechowania. Do możliwych wariantów tego wyrażenia mogą należeć m.in.: *tam to gdzieś; tam to w nosie / мне плевать; мне фиолетово / kaślu na to; je mi to fuk*. Naturalnie, każda z przytoczonych fraz może mieć przypisaną inną konotację (stopniującą intensywność emocjonalnego dystansu), jednakże wyjście od jednej jednostki w kierunku jej wielowyzrazowych odpowiedników zauważalnie pozwala rozszerzyć repertuar fatyzmów i emotywów języka mówionego. Odnalezienie synonimów dla zebranych przez nas uprzednio jednostek umożliwiają przede wszystkim językowe źródła internetowe, takie jak słowniki jednojęzyczne (zawierające frazy wraz z synonimami) czy strony bezpośrednio poświęcone zagadnieniu synonimii. Podejmijmy w tym miejscu próbę zwięzłego przedstawienia możliwości wykorzystania określonych źródeł polskich, rosyjskich oraz czeskich w celu identyfikacji odpowiedników synonimicznych, co umożliwi następnie analizę ich znaczenia, stopnia odpowiedniości względem jednostki wyjściowej oraz ich frekwencji.

Synonimy polskie. Wykorzystanie stron dobrysloownik.pl, sjp.pwn.pl i synonim.net

W celu syntetycznego przedstawienia zastosowania omawianej metody oraz oceny jej przydatności do wyodrębniania wielowyzrazowych jednostek języka mówionego, w analizie zostaną uwzględnione następujące internetowe źródła internetowe:

1. **dobrysłownik.pl** - Projekt, zapoczątkowany przez Artura Czesaka, Łukasza Szalkiewicza i Sebastiana Żurowskiego, który można scharakteryzować jako nowoczesny typ polskiego słownika internetowego, dostępny w wersji online od 2014 roku. Nowatorską cechą tej bazy jest dobieranie haseł zgodnie z potrzebami użytkowników, co nadaje mu charakter „słownika na życzenie”. Należy wspomnieć, że zespół korzysta również z różnych źródeł językowych, w tym NKJP, łącząc jednocześnie słownik, poradnię językową i kompendium zasad poprawnej polszczyzny. Można go określić jako hipertekstowy, poprawnościowy i oparty na rzeczywistym uzusie językowym, co oznacza, że informacje są na bieżąco aktualizowane.

2. **sjp.pwn.pl** (*Słownik języka polskiego PWN*) – *SJP PWN* jest opracowaniem jednojęzycznym, uwzględniającym szeroki zbiór definicji, synonimów, odmian gramatycznych oraz przykładów użycia i ma na celu upowszechnianie poprawności językowej. O staranność każdego artykułu hasłowego dba zespół redaktorów i specjalistów językoznawstwa, a jego zawartość podlega stałej aktualizacji, która umożliwia regularną obserwację zachodzących zmian w polszczyźnie. Użytkownikowi zostaje również udostępniona poradnia językowa, w której można skonsultować się z ekspertami w kwestii nurtującego zagadnienia językowego. Istotna jest również obecność dużej bazy słownikowej synonimów i antonimów.

3. **synonim.net** – strona ta zawiera popularny internetowy słownik synonimów i wyrazów bliskoznacznych. Prosty i intuicyjny interfejs narzędzia umożliwia szybkie wpisanie poszukiwanego słowa i uzyskanie wyników bez zbędnych komplikacji, co umożliwia obszernie zgromadzona baza danych, zawierająca również przykłady antonimów. Należy podkreślić, iż strona nie dostarcza wyczerpujących informacji dotyczących konkretnego wyrazu (np. gramatycznych czy kontekstowych), co czyni ją skierowaną raczej do użytkowników poszukujących podstawowych informacji o słowach bez zaawansowanych funkcji lingwistycznych.

Bazy te będą stanowiły podstawowe narzędzie wykorzystywane podczas metody synonimicznej, polegającej na:

- 1) wykorzystaniu wybranych uprzednio zgromadzonych fatyzmów i emotywów,
- 2) odnalezieniu przy pomocy wskazanych źródeł jednostek synonimicznych,
- 3) sprawdzeniu ich odpowiedniości semantycznej i stopnia rozpowszechnienia w zasobach korpusowych (NKJP, KWJP, MoncoPL).

W celach egzemplifikacyjnych posłużymy się najpierw wybranym fatyzmem i emotywarem, których analizy będą przedstawione we wskazanej kolejności: *moim zdaniem*; *na sto procent*. Następny etap analizy polega na wprowadzeniu danego fatyzmu i emotywu do baz

danych synonimów w celu zidentyfikowania odpowiadających im zbliżonych jednostek wielowyrazowych, a uzyskane wyniki badania przedstawiają się w sposób następujący:

<i>Moim zdaniem...</i>	
ŹRÓDŁO	WYSZUKANE SYNONIMY
<i>dobryslownik.pl</i>	Nie odnaleziono
<i>sjp.pwn.pl</i>	według mnie; w moim odczuciu; w moim odbiorze; na mój rozum; w moim przekonaniu; po mojemu
<i>synonim.net</i>	myślę, że; na mój gust; w mojej opinii; w moim odczuciu; w moim przekonaniu; według mnie

Powszechnie używane wyrażenie *moim zdaniem* służy do wprowadzenia subiektywnej opinii i punktu widzenia mówiącego. Jest ono odnotowane w zasobach korpusowych w liczbie 189 539 poświadczeń. Pomimo nierejestrowania przez źródło *dobryslownik.pl* jednostek synonimicznych dla wskazanej jednostki, jego przydatność w procesie uzyskiwania podobnych komunikatów nie budzi wątpliwości, co zostanie potwierdzone w późniejszej części analizy. Wskazany fatyzm jest używany przede wszystkim w sytuacji subiektywnego podkreślenia przez mówiącego tego, że to, co mówi w danym momencie, jest jego osobistą opinią, np.

Grzegorz: **Moim zdaniem**, Związek popełnił błąd, że nie patrzył na jakość, ale na ilość. Do "Solidarności" przystąpiło wielu prowokatorów, którzy organizowali strajki...

Źródło *dobryslownik.pl* nie odnotowało synonimicznego odpowiednika dla *moim zdaniem*, jednak pozostałe zasoby internetowe dostarczają wielu ekwiwalentów, spośród których znaczna część wykazuje wzajemną zbieżność. Nie ulega wątpliwości fakt, że wszystkie z wyodrębnionych jednostek pełnią identyczną funkcję wyrażania własnej opinii, jednak można wśród nich zaobserwować różnice dotyczące m.in.: rejestru językowego (np. potoczny – *po mojemu*) czy nacechowania emocjonalnego (np. osobiste – *na mój gust*). Poniżej zaprezentowano przykłady zastosowania wyekscerpowanych fatyzmów synonimicznych w kontekście zdaniowym, które ukazują ich funkcję komunikacyjną oraz zróżnicowania stylistyczno-pragmatyczne:

FATYZM	PRZYKŁAD W ZDANIU
<i>według mnie</i>	<i>To prawda, można się tym zarazić, ale jeśli postępujesz rozumnie, to się nie zarazisz, a rzecz jest do wyleczenia. Istnieją też schroniska dla zwierząt. Jednym słowem, według mnie twój ojciec robił ci siano z mózgu.</i>
<i>w moim odczuciu</i>	<i>W moim odczuciu piękno to jest również szlachetność, dobroć, uczciwość, wierność - cały zespół zjawisk idealnych. Piękno to właśnie olśnienie, zachwyt, miłość od pierwszego wejrzenia [...].</i>
<i>w moim odbiorze</i>	<i>W pierwszym odruchu skasowałam ten list, pełen (w moim odbiorze) złości i nienawiści. Potem zdałam sobie sprawę, że chyba chciałabym „kasować” ludzi, którzy reprezentują ten sposób myślenia.</i>
<i>na mój rozum</i>	<i>- Czy nie za wiele się pan domyślasz? Na mój rozum wszystkie agenciaki lubią fundować sobie kolacyjki za państwowe pieniądze.</i>
<i>w moim przekonaniu</i>	<i>Szczepan siadywał w wolne popołudnia na skórzanej kanapce pod oknem i grywał na niej, w moim przekonaniu, przepięknie. Były to siarczyste oberki i polki, wolne walce, albo różne piosenki [...].</i>
<i>po mojemu</i>	<i>- Tom nie powiedział? Kulawa mocno była, ledwie co na nogę stawiała. W bitce jeszcze, jeszcze, ale potem, jak gorączka bojowa opadła, to się siekierką podpierać musiała. Po mojemu to dlatego Hanusa zabrała.</i>
<i>myślę, że</i>	<i>Chciałabym go spotkać, bo mimo że tak mnie potraktował, jest mi go żal. On musi być bardzo samotny i nieszczęśliwy. Myślę, że jakaś kobieta musiała mu zrobić straszną krzywdę.</i>
<i>na mój gust</i>	<i>- Za chuda jesteś jak na mój gust - rzekł - ale ciekawa z ciebie dziewczucha, nie można powiedzieć...</i>
<i>w mojej opinii</i>	<i>W mojej opinii, obaj profesorowie wadliwie interpretują kluczowe terminy: „koegzystencja” i „odprężenie”.</i>

Wszystkie z przedstawionych powyżej fatyzmów służą bez wątpienia do wyrażenia osobistych opinii, odczuć, przekonań czy oceny sytuacji przez mówiącego. Jak można zauważyć, powyższe frazy znajdują swoje zastosowanie w różnorodnych kontekstach – od wypowiedzi oficjalnych, przez potoczne rozmowy, aż po opisy bardziej literackie. Interesującym przykładem wyekscerpowanego fatyzmu jest wyrażenie *po mojemu*, którego przytoczone otoczenie tekstowe prawdopodobnie wskazuje na częstsze używanie danego wyrażenia w mowie silnie potocznej. Należy podkreślić, iż analiza pozostałych poświadczeń fatyzmów pozwala stwierdzić, że wyrażenia te charakteryzują się zróżnicowaną pozycją w strukturze zdania, co oznacza, że mogą być lokowane na jego początku (pełniąc funkcję

wprowadzającą) wewnątrz (gdzie stanowią element rozwijający myśl) lub na końcu (gdzie podsumowują lub treść wypowiedzi). Poniżej przedstawiono odnotowaną w korpusach NKJP, KWJP i MoncoPL łączną liczbę poświadczeń dla każdego z badanych fatyzmów:

FATYZM	FREKWENCJA
<i>według mnie</i>	43 204 poświadczenia
<i>w moim odczuciu</i>	9 247 poświadczeń
<i>w moim odbiorze</i>	118 poświadczeń
<i>na mój rozum</i>	132 poświadczenia
<i>w moim przekonaniu</i>	25 243 poświadczenia
<i>po mojemu</i>	1 478 poświadczeń
<i>myślę, że</i>	53 206 poświadczeń
<i>na mój gust</i>	947 poświadczeń
<i>w mojej opinii</i>	10 466 poświadczeń

Powyższe dane pozwalają wysnuć wniosek, iż najbardziej rozpowszechnionym środkiem wyrażania swojej opinii jest fatyzm wprowadzający *myślę, że*, który wydaje się wykazywać najwyższy stopień neutralności spośród wyodrębnionych jednostek fatycznych. Stosunkowo niska liczba poświadczeń dla frazy *na mój rozum* może wynikać z jej dominującego występowania w rejestrze mówionym oraz w języku potocznym, które – jak wiadomo – są stosunkowo rzadko rejestrowane w dostępnych korpusach językowych. Należy również podkreślić, iż wskazana liczba 1 478 poświadczeń dla jednostki *po mojemu* nie uwzględnia wszystkich rzeczywistych użyc danego fatyzmu w analizowanym przez nas znaczeniu, a to z uwagi na to, że analizowany wielowyrzowiec jest używany również w innym znaczeniu, jak np.

Tarta z jabłkami po mojemu: Ciasto kruche lub francuskie, schłodzone. 2-3 czerwone jabłka, cukier kryształ. Jabłka umyć, osuszyć.

Analiza powyższych danych sugeruje wyraźne różnice w stopniu rozpowszechnienia danych fraz, przy czym najwyższą frekwencję występowania można przypisać prostym i uniwersalnym formom, podczas gdy konstrukcje o bardziej złożonym charakterze występują zdecydowanie rzadziej, przede wszystkim w określonych rejestrach językowych lub specyficznych kontekstach komunikacyjnych. Poniżej zaprezentowano krótki przykład

pozostałych wielowyrazowych fatyzmów, które zostały wyodrębnione metodą synonimiczną we wskazanych przez nas źródłach:

FATYZM	dobryslownik.pl	sjp.pwn.pl	synonim.net
<i>mało tego</i>	<i>nie dość na tym; powiem więcej; nawet lepiej; w dodatku</i>	<i>Nie odnaleziono</i>	<i>co więcej; na dodatek; na domiar tego; prócz tego; do tego</i>
<i>innymi słowy</i>	<i>nie odnaleziono</i>	<i>inaczej mówiąc; to jest; to znaczy</i>	<i>inaczej mówiąc; to jest; innymi słowami; oznacza to</i>
<i>nawiasem mówiąc</i>	<i>a tak poza tym; przy okazji; przy tym</i>	<i>notabene</i>	<i>poza tym; oprócz tego; do tego; właściwie to</i>
<i>Z punktu widzenia</i>	<i>ze stanowiska</i>	<i>pod względem; z perspektywy; w aspekcie</i>	<i>pod kątem; z perspektywy; w rozumieniu</i>

Od analizy jednostek fatycznych przejdźmy zatem do zbadania emotywu *na sto procent*, któremu można przypisać następującą definicję: ‘wyrażenie używane dla podkreślenia pewności mówiącego co do prawdziwości informacji, wydarzenia lub rezultatu. Pełni przede wszystkim funkcję emfatyczną i intensyfikującą w komunikacji potocznej, gdzie metaforyczna liczba sto wyraża pełnię przekonania’ (11 562 konkordancji).

Zadanie tej jednostki ematywnej do wyszukiwarek zbiorów synonimów przyniosło następujące rezultaty, które uwzględniają różnorodne wyrażenia bliskoznaczne i istnieje prawdopodobieństwo, że są wykorzystywane w analogicznych kontekstach komunikacyjnych:

<i>Na sto procent</i>	
ŹRÓDŁO	WYSZUKANE SYNONIMY
<i>dobryslownik.pl</i>	<i>na bank; na pewno; nie ulega wątpliwości</i>
<i>sjp.pwn.pl</i>	<i>bez dwóch zdań; jak w banku; bez wątpienia</i>
<i>synonim.net</i>	<i>z całą pewnością; bez dwóch zdań</i>

Zgodnie z założeniami, wszystkie z wyekscerpowanych jednostek pełnią podobne funkcje w wypowiedzi (podkreślenie pewności mówiącego w kwestii prawdziwości określonej informacji), choć porównywalnie do analizowanego wcześniej fatyzmu charakteryzują się

różnym nacechowaniem stylistycznym. Do wyrażen potocznych możemy odnieść jednostki, takie jak *na bank*, *bez dwóch zdań*, *jak w banku*.

Neutralnością stylistyczną cechują się fatyzmy *na pewno*, *nie ulega wątpliwości*, *bez wątpienia*, *z całą pewnością*. Należy podkreślić, iż każda z przedstawionych jednostek pełni funkcję środka modulującego pewność mówiącego w wypowiedzi i używana jest we własnym, specyficznym kontekście. Pragmatyka tych jednostek realizuje się w ich funkcjach komunikacyjnych, które odzwierciedlają zróżnicowanie stopnia pewności oraz nacechowania stylistycznego. Poniżej przedstawiono przykłady użycia wyekscerpowanych emotywów synonimicznych w kontekstach zdaniowych:

EMOTYW	PRZYKŁAD W ZDANIU
<i>na bank</i>	<i>Już by tu byli, co nie? – Na bank – powiedział Rudzielec.</i>
<i>na pewno</i>	<i>- No to cześć! - Skręciłem w stronę Jarosława. - Na pewno się kiedyś spotkamy. – Na pewno.</i>
<i>nie ulega wątpliwości</i>	<i>- Stwierdził pan tożsamość? - Tak, to ten sam człowiek - powiedział policjant. – To nie ulega wątpliwości. To Horst Kuschka.</i>
<i>bez dwóch zdań</i>	<i>– To jak? Starość nie nadaje się do opisu zmartwychwstania i wiecznego życia? Jest w historii zbawienia tematem tabu? – zagadnął Kolumb. - Bez dwóch zdań!</i>
<i>jak w banku</i>	<i>- Zainwestuje w ciebie. Nie zawieź mnie. – Jak w banku, Pułkowniku.</i>
<i>bez wątpienia</i>	<i>Jurij Łużkow lepiej się nadawał na co dzień... – Bez wątpienia. On był członkiem tego aparatu. Rozumiał go, mógł z nim pracować.</i>
<i>z całą pewnością</i>	<i>- Ale czy to jest udowodnione? – Z całą pewnością. Przeprowadzano zamachy na Elżbietę, które w filmie pokazano.</i>

Przykłady zdaniowe dla wszystkich powyżej przedstawionych emotywów zostały dobrane w taki sposób, aby ukazać ich główną funkcję, czyli funkcję reakcyjną komunikatu umieszczanego na początku wypowiedzi, która stanowi znaczną część zgromadzonych poświadczeń. Niemniej jednak wśród udostępnionych przez korpusy listach konkordancyjnych uwidacznia się silna tendencja wskazanych jednostek do pełnienia funkcji rozwijającej swoją wypowiedź (jako wtrącenie) oraz wzmacniającej przekonanie mówiącego względem wypowiadanych treści. Najbardziej problematycznym dla analizy emotywem z powyższej tabeli jest wyrażenie *na bank*, które w poświadczeniach najczęściej pojawia się w kontekście dosłownym, w towarzystwie kolokatów takich jak *napad* czy *skok*, co wynika z dużej liczby

tekstów opisujących tego rodzaju wydarzenia, np. *Tym, którzy dopiero włączyli radio, przypominamy, że dwaj z prawdopodobnie trzech sprawców dzisiejszego zbrojnego **napadu na bank** w Przywarach zostali ujęci dzięki pomocy mieszkańców*. Łączna liczba odnotowanych w korpusach NKJP, KWJP i MoncoPL poświadczeń dla każdego z wyekscerpowanych emotywów jest następująca:

EMOTYW	FREKWENCJA
<i>na bank</i>	23 215 poświadczeń
<i>na pewno</i>	1 144 192 poświadczenia
<i>nie ulega wątpliwości</i>	26 315 poświadczeń
<i>bez dwóch zdań</i>	7 905 poświadczeń
<i>jak w banku</i>	2 642 poświadczenia
<i>bez wątpienia</i>	107 884 poświadczeń
<i>z całą pewnością</i>	69 742 poświadczenia

Na pierwszy plan dają się zauważyć ogromne różnice w liczbie odnotowań konkretnych jednostek emotywnych przez poszczególne korpusy, co wynika ze wspomnianego faktu zastosowania niektórych z danych wyrażeń w zupełnie odmiennych kontekstach, niż wyrażenia pewności mówiącego:

1. *Na bank* – Obserwacja ponad tysiąca poświadczeń wskazała na dominujące użycie danej jednostki w znaczeniu napadu lub skoku na bank, co sprawia, że oszacowanie rzeczywistej frekwencji emotywu w interesującym nas znaczeniu wymaga każdorazowej weryfikacji poszczególnych konkordancji w korpusie językowym.

2. *Na pewno* – w poświadczeniach tekstowych często współwystępuje z przeczeniem *nie* (*nie na pewno*), co jest antonimem w stosunku do pierwotnego znaczenia, będącego przedmiotem naszej analizy.

Naturalnie, w procesie dalszych badań zasadnym byłoby wskazanie rzeczywistej i przybliżonej liczby poświadczeń dla każdego ze wskazanych emotywów, niemniej jednak ich szerokie zastosowanie zarówno w tekstach pisanych, jak i języku mówionym jest oczywiste. Warto jednak zaznaczyć, że ich szerokie użycie zarówno tekstach pisanych, jak i w języku mówionym nie pozostawia wątpliwości. Poniżej zaprezentowano krótki przykład pozostałych wielowyrzawowych emotywów, które zostały wyodrębnione metodą synonimiczną we wskazanych przez nas źródłach:

EMOTYW	dobryslownik.pl	sjp.pwn.pl	synonim.net
<i>w żadnym razie</i>	<i>w ogóle; w żadnym wypadku</i>	<i>w żadnym wypadku; pod żadnym pozorem; po moim trupie</i>	<i>pod żadnym pozorem; wręcz przeciwnie; w żaden sposób; w ogóle; gdzież tam</i>
<i>nic podobnego</i>	<i>skądże znowu</i>	<i>wcale nie; dupa jasiu; właśnie, że nie</i>	<i>wcale nie; figa z makiem; właśnie, że nie</i>
<i>bez różnicy</i>	<i>Nie odnaleziono</i>	<i>ganc egal; mniejsza o to; bez znaczenia</i>	<i>jeden diabeł; wszystko jedno; wsio ryba</i>
<i>bulka z masłem</i>	<i>kaszka z mlekiem</i>	<i>małe piwo; kaszka z mlekiem</i>	<i>nic trudnego; łatwa rzecz; nic prostszego</i>

Synonimy rosyjskie. Wykorzystanie stron kartaslov.ru, synonymonline.ru i sinonim.org

W celu wyodrębnienia i analizy wielowyrazowych jednostek synonimicznych dla wskazanego rosyjskiego fatyzmu i emotywu posłużymy się następującymi źródłami internetowymi, wśród których część z nich stanowi nie tylko zbiór synonimów czy wyrażen bliskoznacznych, ale można też je traktować jako obszerne słowniki jednojęzyczne:

1. kartaslov.ru – rozbudowany serwis językowy, oferujący użytkownikom przede wszystkim wyszukiwanie wyrazów w słownikach synonimów, antonimów, homonimów czy frazeologizmów. Do udostępnionych narzędzi należy również funkcja analizy tekstu, która pozwala na identyfikację jego cech stylistycznych, zakwalifikowania jednostki do określonego rejestru językowego czy zbadanie struktury składniowej wskazanego zdania. Strona działa w oparciu o zaawansowane algorytmy przetwarzania języka naturalnego, co gwarantuje wysoką dokładność wyników, a prosty w obsłudze interfejs serwisu zwiększa jego popularność zarówno wśród językoznawców, jak i zainteresowanych językiem rosyjskim.

2. synonymonline.ru – Działająca od 2012 roku baza synonimów i powiązanych wyrazów dla języka rosyjskiego, została stworzona głównie z myślą o osobach poszukujących różnorodności językowej, takich jak pisarze, redaktorzy czy studenci. Użytkownik poprzez zadanie zapytania otrzymuje listę synonimów, wyrazów bliskoznacznych czy antonimów w otoczeniach tekstowych, których skuteczna ekscerpcja jest możliwa dzięki bogatej bazie danych językowych, wbudowanych w dane narzędzie. Zyskała wysokie rozpowszechnienie wśród

nieprofesjonalistów czy uczniów, którzy poprzez prostą obsługę strony mogą wyodrębnić potrzebne wyrazy w celu wyeliminowania powtórzeń podczas tworzenia tekstu autorskiego.

3. *sinonim.org* – Baza danych językowych, zawierająca ponad 50 tysięcy słów i umożliwiająca użytkownikom szybkie wyszukiwanie synonimów oraz ich znaczeń w różnych kontekstach. Dla każdej jednostki strona oferuje praktyczne przykłady użycia wyrazów w zdaniach, co zdecydowanie ułatwia zrozumienie różnic stylistycznych między konkretnymi jednostkami i wybór najbardziej odpowiedniego słowa w konkretnej sytuacji. Jedną z największych zalet tego narzędzia jest uwzględnianie aktualnych tendencji językowych, dzięki opcji zgłaszania nowych słów i wyrażen przez samych użytkowników. W odróżnieniu od dostępnych polskich internetowych baz synonimów, które zakładają wyszukiwanie głównie jednostek jednowyrazowych, wskazane źródła rosyjskie w polu wyszukiwania informują o możliwości zadania zapytania w postaci całej frazy. Analogiczną metodę synonimiczną przeprowadzimy na fatyzmie *честно говоря* i emotywie *не может быть*, dla których ustalone zostaną jednostki synonimiczne i ich frekwencja w zbiorach tekstowych NKJR, rosyjskiej części InterCorp v16ud i korpusie Araneum Russicum Maius. Zaproponowane przez narzędzia frazy synonimiczne dla wybranego przez nas fatyzmu są następujące:

<i>Честно говоря...</i>	
źródło	WYSZUKANE SYNONIMY
<i>kartaslov.ru</i>	<i>По правде говоря; мягко говоря; собственно говоря</i>
<i>slovaronline.com</i>	<i>по правде говоря; по совести сказать</i>
<i>sinonim.org</i>	<i>по правде сказать; честно сказать; если быть честным; если начистоту; на самом деле</i>

Wyrażenie *честно говоря* należy traktować jako zwrot wprowadzający, który podkreśla szczerłość i otwartość treści wypowiedzi ze strony mówiącego, a łączna liczba oszacowanych dla niego poświadczeń wyniosła blisko 19 tysięcy konkordancji we wskazanych zbiorach. Zdecydowana większość uwzględnionych w tabeli fatyzmów rzeczywiście może pełnić funkcję synonimiczną dla fatyzmu *честно говоря*, co pozwala na przypisanie im ogólnej definicji:

‘Wyrażenie używane dla podkreślenia szczerości wypowiedzi mówiącego i podkreślenia jego prawdziwej opinii w omawianej kwestii’.

Niemniej jednak, część fatyzmów zaproponowanych przez narzędzia znacząco odbiega od przedstawionej definicji pod względem semantycznym, co sugeruje, że mogłyby one być użyte w odmiennych intencjach oraz w innych kontekstach komunikacyjnych:

1. *Мягко говоря* – zwrot wprowadzający złagodzenie osądu mówiącego, który pozwala uniknąć bezpośredniej krytyki i w sposób delikatny zasugerować, że rzeczywisty obraz sytuacji jest trudniejszy, niż zostało to opisane, np. *Все понимали, что политика в этой истории играет, мягко говоря, не последнюю роль.*

2. *Собственно говоря* – fraza używana w celu jasnego wyrażenia istoty sprawy w formie przedstawionej przez mówiącego, najczęściej w formie wtrącenia, np. *Правда, на время ее приездов Сергею с Таней и Женькой приходилось полностью перебираться в мастерскую, на чем Полуэктова, собственно говоря, не настаивала: ревность была ей не знакома.*

3. *На самом деле* – wyrażenie stosowane do podkreślenia prawdziwego stanu rzeczy, przeciwstawienia się pozorom lub do wprowadzenia dodatkowego wyjaśnienia (możliwy synonim *собственно говоря*), np. *Вы можете подумать, что вас это не касается и даже рассмеяться, и сказать «какая разница, сколько водорослей в водоеме»! Но на самом деле все гораздо серьезнее.*

Wyekscerpowanie przez te źródła internetowe danych jednostek można z jednej strony potraktować jako błąd, a z drugiej – wykorzystać uzyskane jednostki do rozbudowywania bazy fatyzmów i kolejnego odnalezienia dla nich wyrażen synonimicznych. Przykłady zastosowania odnalezionych fatyzmów synonimicznych i tych o odmiennej semantyce w kontekstach zdaniowych zostały przedstawione poniżej:

FATYZM	ПРИКЛАД В ЗДАНИУ
по правде говоря	- В «47 ронинов» вы держитесь как Бастер Китон, несмотря на то что играете жуткого злодея. - Спасибо за комплимент. Я, по правде говоря , нигде актерскому мастерству не учился.
мягко говоря	Подойти к нему можно, но говорить — только в ее присутствии. Нина — человек, мягко говоря , непростой, довольно субъективный в своих оценках и рабски преданный Светланову.
собственно говоря	Где купидоны, амурь и прочие статисты этого захватывающего шоу? Где, наконец, букет цветов за рубль тридцать?!.. Собственно говоря , я даже не знаю, что такое любовь.

по совести сказать	<i>Но если ты уедешь, то мне гораздо труднее будет сделать для тебя что-нибудь. По совести сказать, я не представляю себе, как это организовать.</i>
по правде сказать	<i>— Это по поводу его нашумевшего романа «Стальные рекорды»? — спросил Валентин Алексеевич. — И о романе, и о нём самом, — кивнул Бочкин. — Да я, по правде сказать, роман этот не читал.</i>
честно сказать	<i>Ну а ограничения на использование нецензурной лексики вас не смущают? — Сложный вопрос. Честно сказать, я не переношу мат в кино, хотя в жизни могу крепко выразиться.</i>
если быть честным	<i>Если быть честным, надо признать, что "мои знакомые" — это те люди, с которыми, по чистой случайности, мне пришлось познакомиться, хотя бы поверхностно пообщаться.</i>
если начистоту	<i>Новый год — новая жизнь. Банально? Но если начистоту, — все мы любим начинать «новую жизнь» с какой-нибудь красивой даты.</i>
на самом деле	<i>На самом деле, если кто-то раньше успешно выполнил определенную работу, вряд ли он повторит этот результат снова.</i>

Observacja poświadczeń tekstowych, w których zostały zarejestrowane wskazane fatyzmy, pozwala stwierdzić, iż przytoczone wyrażenia mogą różnić się od siebie, np. nacechowaniem stylistycznym:

Oficjalne: *По совести сказать; по правде говоря.*

Potoczne: *Честно сказать; если начистоту; если быть честным.*

Nie ulega wątpliwości fakt, że wyrażenie *по совести сказать* charakteryzuje się większą oficjalnością niż synonim *по правде говоря*. Dlatego też, pomimo ich niewątpliwego podobieństwa semantycznego, każda z jednostek zostanie użyta w nieco odmiennych sytuacjach komunikacyjnej, przy innych intencjach autora wypowiedzi i w inaczej określonym celu. Frekwencja poszczególnych jednostek w zbiorach NKJR, rosyjskiej części InterCorp v16ud i korpusu Araneum Russicum Maius (przedstawiająca sumę wszystkich poświadczeń) kształtuje się następująco:

FATYZM	FREKWENCJA
<i>по правде говоря</i>	3 286 poświadczeń
<i>мягко говоря</i>	10 530 poświadczeń
<i>собственно говоря</i>	9 488 poświadczeń

<i>по совести сказать</i>	148 poświadczeń
<i>по правде сказать</i>	1 834 poświadczenia
<i>честно сказать</i>	2 062 poświadczenia
<i>если быть честным</i>	401 poświadczeń
<i>если начистоту</i>	87 poświadczeń
<i>на самом деле</i>	183 078 poświadczeń

Na podstawie powyższych danych można wnioskować, iż większość z wyodrębnionych fatyzmów cechuje wysoki stopień użycia wśród użytkowników języka rosyjskiego, co wskazuje na ich uniwersalność i obecność w różnych rejestrach językowych. Natomiast rzadziej występujące fatyzmy, takie jak *по совести сказать* czy *если начистоту*, pojawiają się głównie w specyficznych kontekstach wymagających podkreślenia szczerości lub wyrazistości opinii. Obserwacja udostępnionych przez korpusy narzędzi pozwala stwierdzić, że wszystkie z przedstawionych fatyzmów pełnią funkcję wprowadzającą do wypowiedzi (na początku zdania) bądź metatekstową, przejawiającą się w ich użyciu jako wtrąceń w strukturze tekstu. Poniżej przedstawiono krótki przegląd pozostałych fatyzmów wielowyrazowych, wyodrębnionych za pomocą metody synonimicznej na podstawie wskazanych źródeł:

FATYZM	<i>kartaslov.ru</i>	<i>slovaronline.com</i>	<i>sinonim.org</i>
<i>Мне кажется</i>	<i>на мой взгляд; мне видится; как я погляжу</i>	<i>с моей точки зрения; по-моему; насколько я могу судить</i>	<i>по моему мнению; здаётся мне; по моим мыслям; я думаю, что</i>
<i>Можно сказать</i>	<i>как говорится; едва ли не; так называемый; так сказать</i>	<i>надо сказать; признать сказать; как бы сказать</i>	<i>я бы сказал; чуть ли не; как говорится; что называется</i>
<i>Короче говоря</i>	<i>одним словом; словом сказать</i>	<i>коротко говоря; словом сказать,</i>	<i>одним словом; в целом; словом сказать</i>
<i>В конце концов</i>	<i>в конечном итоге; в результате; как-никак</i>	<i>в конечном счёте</i>	<i>в конечном итоге; рано или поздно; в результате</i>

Kolejna część zostanie poświęcona próbie znalezienia jednostek synonimicznych dla emotywu *не может быть*, któremu można przypisać następującą definicję:

‘zwrot potoczny dla wyrażenia zdziwienia bądź niedowierzania wobec sytuacji, informacji lub faktu, które wydają się sprzeczne z dotychczasową wiedzą czy oczekiwaniami mówiącego’.

Ustalenie przybliżonej liczby poświadczeń dla wskazanej jednostki jest problematyczne z uwagi na to, iż dana konstrukcja często występuje w tekstach w znaczeniu dosłownym, tj. zaprzeczeniu, które wyraża brak możliwości zrealizowania określonego działania. Analiza tej jednostki emotywniej za pomocą wskazanych źródeł internetowych do wyszukiwania synonimów pozwoliła zidentyfikować szereg wyrażen bliskoznacznych, które prawdopodobnie mogą być użyte w podobnych sytuacjach komunikacyjnych:

<i>Не может быть</i>	
źródło	WYSZUKANE SYNONIMY
<i>kartaslov.ru</i>	<i>во сне не снилось; не верится; не укладывается в голове</i>
<i>slovaronline.com</i>	<i>слыханное ли дело; не верится</i>
<i>sinonim.org</i>	<i>не верится; вот это да; вот это приколы; вот это новость; взятки гладки</i>

Wspólną cechą większości z wyekscerpowanych jednostek jest ich analogiczne użycie w sytuacji, gdy mówiący chce wyrazić swoje zaskoczenie i zdumienie z powodu przekazanej mu informacji. Najczęstszym miejscem analizowanych wyrażen jest ich lokalizacja na początku wypowiedzi, gdzie pełnią funkcję gotowych replik służących natychmiastowemu wyrażeniu emocji. Ponadto, niektóre frazy mogą sugerować trudność w zrozumieniu lub zaakceptowaniu otrzymanej wiadomości poprzez elementy metaforyczne, np. *не укладывается в голове; во сне не снилось*. Z kolei wyekscerpowana jednostka *слыханное ли дело*, pomimo podobnej funkcji wyrażania zdziwienia, charakteryzuje się bardziej archaicznym i retorycznym stylem, zawiera element osądu i zazwyczaj pełni rolę wtrącenia w wypowiedzi. Potocznością odznaczają się jednostki *вот это да!; Вот это приколы!; Вот это новость!*, przeważnie łączące element wyrażenia zdziwienia z pozytywną reakcją ze strony mówiącego. Wartym uwagi jest wyodrębniony emotywny wyraz *взятки гладки*, którego nie można zaliczyć do fraz synonimicznych dla jednostki wyjściowej *не может быть* z powodu jego odmiennego znaczenia: ‘używane dla opisu osoby, która nie ponosi żadnej odpowiedzialności za kogokolwiek lub coś, od której nie można niczego oczekiwać ani otrzymać’ (Мокиенко, Никитина 2007: 85), np. *Сиди и помалкивай: мы — деревенские, люди темные, с нас*

взятки гладки. *Спать будем по очереди*. Dla tej jednostki, podobnie jak dla pozostałych, wyodrębniono przykłady użycia w zdaniach, które zostały zaprezentowane poniżej w celu ilustracji ich funkcji i kontekstów zastosowania:

ЕМОТЫВ	ПРИКЛАД В ЗДАНИУ
во сне не снилось	— <i>Теперь я вижу, что вы и в самом деле станете военными. Какая прелесть! Две дочери — и обе генералы! Мне такое счастье и во сне не снилось!</i>
не верится	— <i>По крайней мере для нее, — согласилась мама. — Анисия Ивановна, как всегда, права. — И крикнула в папину сторону: — Просто не верится, что ты ее родственник!</i>
не укладывается в голове	<i>Причём лично ко мне никто не обращался. За 17 лет моей практики такое впервые. Это не укладывается в голове! Пятнадцатый век! Как такое возможно в наше время?</i>
слыханное ли дело	<i>Вон там за кустом лис с огненным хвостом прочертил, — а нас не спускают! Шарахнулась тут свита, завертелись охотнички... Слыханное ли дело, чтобы людям молчать, псам разговаривать?</i>
вот это да	— <i>А что такое вот это «Б.П.»? — Знаешь, а ты ведь угадал — действительно весенняя. Вообще первоначально эта штука называлась «Зелёные кеды». — Вот это да!</i>
вот это прикол	<i>Особенно мне понравились его красные труселя! Вот это прикол! Абрамович аж рот открыл от удивления!</i>
вот это новость	— <i>Я беременна, — беззвучно повторяет Тамара. Вот это новость так новость! Я в шоке; я припадаю к спасительному устою колонны, чтобы удержаться на ногах.</i>
взятки гладки	— <i>Ой, девочки, мне бы ваши годы! — мечтательно ответила Марина Михайловна. — Хотя нет, не хочу. Сейчас я могу позволить себе любую глупость. Ну, назовут чокнутой старухой — и все, с меня взятки гладки.</i>

Większość z przedstawionych wyrażen można odnieść do rejestru języka potocznego, przy pełnieniu funkcji silnie ekspresywnych i stanowiących werbalny środek wyrażania zdziwienia, niedowierzania, aprobaty, dezaprobaty lub nawet szoku. Zauważalne są różnice dotyczące nacechowania emocjonalnego, które mogą wyrażać odbiór pozytywny (*вот это да*), ton ironiczny (*вот это новость*) czy oburzenie i krytykę osoby mówiącej (*слыханное ли дело*). Częstym zjawiskiem są współwystępujące z określoną jednostką elementy, które

sprawiają, że mamy w tekście do czynienia z wariantem konkretnego wielowyrazowego emotywu, jak np.

не верится → **Просто** не верится!; *Не укладывается в голове* → <кому> не укладывается в голове.

Kolejna część analizy zawiera się w określeniu łącznej liczby poświadczeń dla każdego z wyekscerpowanych jednostek emotywnych w uprzednio wskazanych zbiorach korpusowych.

EMOTYW	FREKWENCJA
<i>во сне не снилось</i>	96 poświadczeń
<i>не верится</i>	5 472 poświadczenia
<i>не укладывается в голове</i>	241 poświadczeń
<i>слыханное ли дело</i>	69 poświadczeń
<i>вот это да</i>	2 892 poświadczenia
<i>вот это приколы</i>	22 poświadczenia
<i>вот это новость</i>	158 poświadczeń
<i>взятки гладки</i>	286 poświadczeń

Obserwacja danych frekwencyjnych wyekscerpowanych jednostek emotywnych ujawnia znaczące różnice w rozpowszechnieniu poszczególnych wyrażeń, co w przypadku niskiej frekwencji jednostek potocznych wynika najprawdopodobniej z ich nieuwzględnienia w zasobach tekstowych wybranych korpusów. Wymienione frazy, takie jak *вот это приколы* czy *вот это новость* z uwagi na ich potoczny charakter i silne nacechowanie emocjonalne, występują głównie w komunikacji ustnej, co może tłumaczyć ich relatywnie niską frekwencję w korpusach językowych opartych na materiałach pisanych, gdzie tego typu jednostki są praktycznie nieobecne. Pozyskana liczba poświadczeń potwierdziła również uprzednio sformułowaną charakterystykę emotywu *слыханное ли дело*, którego archaiczność oraz spadek częstości użycia w języku współczesnych użytkowników rosyjskiego skutkują jego ograniczonym występowaniem i rzadkim odnotowaniem w analizowanych tekstach korpusowych. Wskazane źródła internetowe zostały wykorzystane do odnalezienia jednostek synonimicznych dla pozostałych emotywów, a rezultaty analizy przedstawiają się następująco:

EMOTYW	<i>kartaslov.ru</i>	<i>slovaronline.com</i>	<i>sinonim.org</i>
чёрт знает	<i>Бог ведает; пёс знает; бог знает</i>	<i>Бог ведает; бог знает; пёс знает</i>	<i>Чёрт возьми; бог знает; пёс знает</i>
вовсе нет	<i>А вот и нет; никак нет; куда там; да вы что!</i>	<i>Да ты что! Какое там! Как бы не так</i>	<i>Совсем нет; что ты; какое там; ничуть не бывало</i>
честное слово	<i>Даю слово; клянусь матерью; право слово</i>	<i>Гадом буду; ей-ей; зуб даю</i>	<i>Сто пудов; без обмана; право слово</i>
никогда в жизни	<i>Ни в жизни; ни разу; ни разу в жизни</i>	<i>Ни за что на свете; через мой труп; ни в жизнь</i>	<i>Ни при каких условиях; ни разу в жизни</i>

Synonimy czeskie. Wykorzystanie stron dobryslovník.cz, slovníkcestiny.cz i esynonyma.cz

Trudność w analogicznym zastosowaniu metody synonimicznej w przypadku źródeł poświęconych językowi czeskiemu wynika z faktu, że dostępne elektroniczne bazy synonimów czeskich w większości nie pozwalają na wyszukiwanie jednostek wielowyrazowych. Wśród najpopularniejszych czeskich platform stworzonych do wyszukiwania odpowiedników znaczeniowych nie znaleźliśmy reproduktów fatycznych ani emotywnych w dostępnych bazach, a były to: *nechybujete.cz*; *slovník-synonym.cz* (abz slovník českých synonym); *synonymus.cz*; *synonyma-online.cz*.

Próby zidentyfikowania innych czeskich źródeł internetowych uwzględniających możliwość wprowadzenia jednostki wielowyrazowej i odnalezienia jej wieloelementowego synonimu również napotkały trudności, jednak w toku badań udało się wskazać pewne źródła, które, choć w ograniczonym zakresie, uwzględniały jednostki wielowyrazowe oraz listę dostępnych dla nich synonimów:

1. *dobryslovník.cz* – jest internetowym słownikiem wielojęzycznym, wyposażonym w interakcyjne narzędzia lingwistyczne do analizy językowej, umożliwiającej odnalezienie definicji, synonimów, antonimów czy przykładów użycia jednostki w zdaniu. Strona zawiera językowe bazy danych dla m.in. języka czeskiego, angielskiego, bułgarskiego, francuskiego czy rosyjskiego. Platforma umożliwia również analizę słów w kontekście, dzięki czemu użytkownik może lepiej zrozumieć ich znaczenie czy zastosowanie w określonej sytuacji komunikacyjnej. Warto podkreślić, że w polu wyszukiwania jest możliwość zadania pytania nie tylko w postaci hasła jednowyrazowego, ale również w formie frazy czy nawet całego

zdania. Regularnie aktualizowane zasoby oraz intuicyjna w obsłudze wyszukiwarka czynią to narzędzie teoretycznie przydatnym w procesie analizy wybranego przez użytkownika języka.

2. *slovníkcestiny.cz* (*Akademický slovník současné češtiny*) - projekt realizowany przez Instytut Języka Czeskiego Akademii Nauk Republiki Czeskiej, którego celem jest dokumentacja i analiza współczesnego języka czeskiego, a także wsparcie podejmowanych badań w obszarze języka czeskiego. Słownik zawiera szczegółowe definicje wyrazów, ich użycie w kontekście, informacje o odmianie gramatycznej oraz charakterystykę stylistyczno-pragmatyczną, a unikalną cechą słownika jest uwzględnianie nie tylko oficjalnych form językowych, ale także wariantów potocznych, co pozwala scharakteryzować niniejszą bazę jako kompleksowe źródło wiedzy o współczesnej czeszczyźnie. Platforma jest regularnie aktualizowana i dostępna bezpłatnie, a wysoka przystępność w obsłudze czyni ją wartościowym narzędziem dla badaczy, studentów oraz pozostałych osób zainteresowanych współczesnymi tendencjami językowymi czeszczyzny.

3. *esynonyma.cz* – jest internetowym słownikiem synonimów języka czeskiego, który umożliwia szybkie i intuicyjne wygenerowanie listy odpowiedników semantycznych dla wskazanej przez użytkownika jednostki. Oprócz synonimów, strona nierzadko oferuje również wskazówki dotyczące niuansów znaczeniowych i stylistycznych czy odnosi się w swoich zasobach do bardziej archaicznych i rzadziej używanych wyrazów, co pozwala na pełny ogląd tendencji rozwojowych języka czeskiego. Platforma pomaga wzbogacić słownictwo, unikać powtórzeń w tekście, rozwijać zdolności językowe czy wspierać nauczanie języka czeskiego, będąc przy tym bezpłatnym narzędziem dla użytkowników na różnych poziomach zaawansowania.

Należy podkreślić, że wskazane bazy internetowe w sposób ograniczony i rozproszony odnotowują niektóre z interesujących nas fatyzmów i emotywów, przy czym najczęściej obserwowano sytuację, w której daną jednostkę rejestrowało jedynie jedno ze źródeł. Dlatego wyniki zastosowania metody synonimicznej w omawianym przypadku zostaną zaprezentowane w sposób bardziej zróżnicowany niż w analizie fatyzmów i emotywów w języku polskim czy rosyjskim. Pierwszym z analizowanych fatyzmów, wyekscerpowanym z bazy *dobryslownik.cz* jest jednostka *tak tedy*, której można przypisać następujące znaczenia: ‘zwrot używany dla sygnalizowania, że mówiący w swej wypowiedzi przechodzi do sedna omawianej sprawy/ dla potwierdzenia, że coś zostało ustalone bądź zrozumiałe/jako środek retoryczny wyrażający chwilę zastanowienia’ (33 990 konkordancji). Poniżej zaprezentowano zaczerpnięte ze wskazanego źródła przykłady odpowiedników synonimicznych, przykłady ich użycia w

zdaniach oraz łączną liczbę poświadczeń dla każdego z nich w oparciu o korpus Syn v8, czeską część InterCorp v16ud oraz zbioru Araneum Bohemicum Maximum:

Tak tedy	
źródło	WYSZUKANE SYNONIMY
dobryslownik.cz	<i>V takovém případě; v tomto pořadí; v tom případě</i>

FATYZM	PRZYKŁAD W ZDANIU
<i>V takovém případě</i>	<i>A jsme nejlepší přátelé. Dobře. A řekněme, když to Rob zkazí a celý večer bude fiasko, tak já ihned přispěchám a zachráním to. - Dobře, v takovém případě klidně.</i>
<i>v tomto pořadí</i>	<i>Když tuto informaci obdrželi samotní členové Pearl Jam, reagovali dvěma otázkami v tomto pořadí: 1. Jak se to stalo? 2. Líbila se písnička lidem?</i>
<i>v tom případě</i>	<i>Jste i vy věčně nespokojená, protivná, náladová, mimořádně negativně naladěná? V tom případě máte přímo na čele napsaný důvod, proč se vám muži obloukem vyhýbají.</i>

FATYZM	FREKWENCJA
<i>V takovém případě</i>	111 532 poświadczenia
<i>v tomto pořadí</i>	4 717 poświadczeń
<i>v tom případě</i>	81 356 poświadczeń

W przypadku wyodrębnionych jednostek *V takovém případě* / *v tom případě* można założyć, iż stanowią one warianty jednego i tego samego re produktu, wykazując wysokie podobieństwo semantyczne względem jednostki wyjściowej *tak tedy*. Dane wyrażenie najczęściej występuje w procesie formułowania wniosków, reagowania na nowe okoliczności lub podejmowania decyzji w oparciu o określone warunki, przy czym zazwyczaj pojawia się na początku zdania jako odpowiedź na uprzednio omówioną sytuację. Ich frekwencja jest wysoka w języku czeskim, przy czym konstrukcja *v takovém případě* jest częściej używana i odnotowywana w analizowanych tekstach korpusowych, co może wskazywać na jej większą wartość komunikacyjną. Z kolei jednostka *v tomto pořadí* została błędnie wskazana przez narzędzie jako synonim do jednostki *tak tedy*, ponieważ jej definicja znacząco odbiega od znaczenia, które stanowi przedmiot analizy: ‘sformułowanie odnoszące się do sekwencji elementów bądź czynności, których kolejność występowania została wcześniej dokładnie

określona'. Znaczna liczba poświadczeń umożliwiła zaobserwowanie odmiennej funkcji oraz semantyki analizowanego wyrażenia, co uzmysławia konieczność każdorazowej weryfikacji proponowanego przez platformę synonimu w celu jego dokładnej oceny i wykluczenia ewentualnych błędów.

Kolejnym z zaprezentowanych fatyzmów odnalezionym w źródle *slovníkcestiny.cz*, jest formuła wprowadzająca *Dost na tom, že...*, dla której ustalona liczba konkordancji we wskazanych korpusach wyniosła 960 i można jej przypisać następującą semantykę: 'zwrot służący do wyrażenia niezadowolenia, irytacji i podkreślenia, że określona sytuacja bądź okoliczności są wystarczającym powodem do czegoś i same w sobie są już wystarczająco problematyczne'. Wyekscerpowany jeden przykład odpowiednika synonimicznego wraz z poświadczeniem tekstowym i liczbą użycy zaprezentowano poniżej:

<i>Dost na tom, že...</i>	
źródło	WYSZUKANE SYNONIMY
<i>slovníkcestiny.cz</i>	<i>stačí, že...</i>

FATYZM	PRZYKŁAD W ZDANIU
<i>Stačí, že...</i>	<i>Snad ho nenapadne, že by se mohl proměnit v letadlo a odstartovat ze špičky majáku!" "Hlavně to před ním neříkejte, " zasmála se teta Fanny. "Stačí, že si pořád hraje na ta svoje auta"</i>

FATYZM	FREKWENCJA
<i>stačí, že...</i>	7 700 poświadczeń

Wskazane źródło zaproponowało ekwiwalent *stačí, že...*, który można uznać za jednostkę o szerszym zakresie użycia, przejawiającą się w kontekstach zarówno pozytywnych, jak i negatywnych. O jego uniwersalności świadczy zdecydowanie wyższa frekwencja użycia, która wskazuje na jego główną funkcję polegającą na podkreślanu, że określony fakt, okoliczność lub sytuacja stanowią wystarczający powód bądź czynnik prowadzący do określonych skutków. W analizowanym przypadku podobieństwo semantyczne zostałyby w pewnym stopniu zachowane, gdyby dane zwroty były stosowane wymiennie w kontekście dyskusji dotyczących problemów lub trudności, przy założeniu, że zachowane zostaną odpowiednie ramy kontekstualne i intencje komunikacyjne nadawcy wypowiedzi.

Ostatnim z badanych fatyzmów zarejestrowanym w źródle *esynonyma.cz*, jest jednostka *v podstatě*, której można przypisać następującą definicję: ‘wyrażenie używane do podkreślenia sedna sprawy, podsumowania informacji bądź wskazania głównego wątku omawianej kwestii’. Dla tego reperytorium wyeksцерpowano następujące możliwości synonimiczne, uzupełnione przykładami ich użycia w zdaniu oraz oszacowaną frekwencją we wskazanych uprzednio korpusach:

<i>V podstatě</i>	
źródło	WYSZUKANE SYNONIMY
<i>esynonyma.cz</i>	<i>ve skutečnosti; v zásadě; v základním principu</i>

FATYZM	PRZYKŁAD W ZDANIU
<i>ve skutečnosti</i>	<i>Na postup čeká roky - ve firmě má nulový vliv, je přesvědčován o opaku, jeho firemní reporty stejně končí v koši pokud se domáhá na manažerech jejich řešení, je brzo z firmy vyhozen protože ve skutečnosti jen přidělová práci manažerům.</i>
<i>v zásadě</i>	<i>Klinicky se používají dva preparáty: Dysport a Botox. Více známý je Botox, ale v zásadě se tyto deriváty botulotoxinu od sebe neliší, snad jen koncentrací roztoku.</i>
<i>v základním principu</i>	<i>To je podle mne směřování, které bychom měli v základním principu podpořit. Zkušenosti ze života nám však ukazují, že dobrým řešením je i zveřejňování smluv na internetu [...].</i>

FATYZM	FREKWENCJA
<i>ve skutečnosti</i>	376 648 poświadczeń
<i>v zásadě</i>	100 429 poświadczeń
<i>v základním principu</i>	62 poświadczenia

Wśród wyodrębnionych fatyzmów największy stopień odpowiedniości semantycznej posiada jednostka o wysokiej frekwencji *v zásadě*, której używa się głównie do wyrażenia ogólnego zarysu danego tematu bądź wskazania pewnych ograniczeń czy zastrzeżeń w określonej sytuacji. Charakteryzujący się minimalną liczbą poświadczeń zwrot *v základním principu* znaczeniowo odpowiada analizowanej jednostce wyjściowej, gdyż analogicznie akcentuje istotę konkretnego zagadnienia w wypowiedzi. Z kolei jednostka fatyczna *ve skutečnosti* służy bardziej do podkreślenia prawdziwego stanu rzeczy i skonstrastowania czegoś z wcześniejszym twierdzeniem, pozorem lub nieprawidłowym założeniem. Jednakże w

niektórych kontekstach może być stosowana wymiennie względem jednostki *v podstatě*, kiedy dochodzi do przedstawienia ogólnych wniosków i głównych okoliczności omawianej kwestii. Poniżej zaprezentowano przykłady pozostałych pojedynczych jednostek fatycznych, które mogą stanowić podstawę do dalszych badań i analiz, odnotowanych w sposób ogólny na podstawie danych zebranych z przedstawionych źródeł internetowych.

FATYZM	<i>dobryslownik.cz, slovníkcestiny.cz, esynonyma.cz</i>
<i>To mi něco připomíná</i>	<i>To mi přijde povědomé; to mi něco říká</i>
<i>Tak či onak</i>	<i>V každém případě; každopádně</i>
<i>A tak dál</i>	<i>A tak podobně; a jiné</i>

Ostatnią częścią analizy w obszarze metody synonimicznej będą wyodrębnione z czeskich źródeł internetowych przykłady jednostek emotywnych, wśród których pierwszym zwrotem wyodrębnionym z bazy *dobryslownik.cz*, jest silnie rozpowszechniona jednostka *není zač*, która pełni funkcję grzecznościowej odpowiedzi na podziękowanie, wskazując przy tym na brak potrzeby odwzajemnienia wdzięczności ze strony rozmówcy. Źródła tekstowe rejestrują emotyw w ponad 11 tysięcy konkordancjach, dla którego odnaleziono poniższe emotywy wraz z ich użyciem w zdaniach i oszacowaną frekwencją:

<i>Není zač</i>	
źródło	WYSZUKANE SYNONIMY
<i>dobryslownik.cz</i>	<i>Rádo se stalo; To nestojí za řeč; Za málo</i>

EMOTYW	PRZYKŁAD W ZDANIU
<i>Rádo se stalo</i>	„Děkuji vám, že jste si na nás udělala čas, Tegan, řekl Strike. Trochu se začervenala a potřásla si rukou napřed s ním a potom s Robin. „ Rádo se stalo ,“ řekla a najednou se zdálo, že se jí nechce odejít.
<i>To nestojí za řeč</i>	"To je maličkost, slečno, "bránil se. "Náhodou... mám zrovna volný čas... stejně chci někam jet, a vůbec - "Mávl v rozpacích rukou. " To nestojí za řeč . Udělám všechno, co chcete."
<i>Za málo</i>	Celé to bylo nanejvýš nepříjemné. Povzdechla si. „Děkuji,“ řekla a William sebou trhl. „ Za málo ,“ vykoktal.

EMOTYW	FREKWENCJA
<i>Rádo se stalo</i>	2 237 poświadczeń
<i>To nestojí za řeč</i>	1 241 poświadczeń
<i>Za málo</i>	9 658 poświadczeń

Pierwszy z zaproponowanych synonimów wielowyrazowych wyrażenia *rádo se stalo* można scharakteryzować jako niemal tożsamy pod względem znaczeniowym z analizowaną jednostką wyjściową, pełniący analogiczną funkcję grzecznościowej odpowiedzi na podziękowanie i przy jednoczesnej sugestii, że osoba udzielająca pomocy uczyniła to z radością i bez oczekiwania odwzajemniania. Z kolei jednostka *to nestojí za řeč* jest używana nie tylko w sytuacji odpowiedzi na niewielką przysługę ze strony mówiącego, ale również w sytuacjach kiedy coś jest uznane za nieistotne, niegodne uwagi lub zbyt mało ważne, aby poświęcać temu osobną dyskusję. Dlatego też w danym przypadku zbieżność semantyczna można ocenić jako zachowaną, ale pod warunkiem użycia danych jednostek w podobnych kontekstach. Ostatni ze wskazanych emotywów *za málo* zachowuje odpowiedniość semantyczną, polegającą na sugerowaniu ze strony mówiącego, że pomoc lub uprzednia czynność była zbyt mało istotna, by wymagała wdzięczności. Jednakże wskazana liczba poświadczeń tekstowych uwzględnia w dużej mierze inne, dosłowne zastosowanie danej jednostki wielowyrazowej w kontekście określonego niedoboru czegoś, np.

Poté se na ulicích sežene vhodný počet bezdomovců. Budou na pozemku pracovat za málo peněz a za ubytování a jídlo.

Kolejnym przykładem emotywu, zarejestrowanym w bazie *slovníkcestiny.cz* jest reprodut *To je detail*, którego łączna liczba poświadczeń wyniosła 503 konkordancje, a obserwacja kontekstów użycia pozwoliła sformułować następującą definicję:

‘Używane jest w sytuacji, kiedy jakaś kwestia jest uznawana za mało istotna lub niezaskługująca na szczególną uwagę w danej dyskusji’.

Wyekscerpowane synonimy wielowyrazowe, a także przykłady użyc w zdaniach i frekwencja przedstawiają się następująco:

<i>To je detail</i>	
źródło	WYSZUKANE SYNONIMY
<i>slovníkcestiny.cz</i>	<i>Je to málo důležité; to je maličkost</i>

EMOTYW	PRZYKŁAD W ZDANIU
<i>Je to málo důležité</i>	<i>Co je ale důležité: Podařilo se vám to vypít. Ať už to chutná divně, nebo ne... je to málo důležité.</i>
<i>to je maličkost</i>	<i>- Ale tyhle hrozně rovnice s více neznámými, to je můj problém. - Ale to je maličkost. Vysvětlím ti princip - a hned to pochopíš.</i>

EMOTYW	FREKWENCJA
<i>Je to málo důležité</i>	1 poświadczenie
<i>to je maličkost</i>	544 poświadczenia

Pomimo faktu, iż pierwsza z wyodrębnionych jednostek *Je to málo důležité* wykazuje znaczną zbieżność semantyczną z jednostką wyjściową i jest używana w celu pominięcia kwestii, które nie mają istotnego wpływu na główny temat rozmowy, jej pojedyncze poświadczenie w trzech wybranych korpusach budzi wątpliwości co do jej rozpowszechnienia w języku czeskim. Ten przykład tekstowy został zarejestrowany w bazie OpenSubtitles, dlatego też istnieje duże prawdopodobieństwo, że fraza została dosłownie przetłumaczona z innego języka, gdzie, pomimo zachowania sensu, forma dosłownego tłumaczenia nie znajduje zastosowania w czeszczyźnie mówionej.

Z kolei emotyw *to je maličkost* jest często używany, zachowuje wymagany stopień odpowiedniości semantycznej, ale należy zauważyć, że może być użyty również w funkcji podziękowań za niewielki gest.

Ostatnie z czeskich internetowych źródeł synonimów, tj. strona *esynonyma.cz*, nie odnotowało żadnego przykładu jednostki emotywnej, w związku z czym poniżej zaprezentowane synonimy dla wskazanych emotywów wyjściowych nie uwzględniają tej bazy:

EMOTYW	<i>dobryslovník.cz, slovníkcestiny.cz</i>
<i>Dobře mu tak</i>	<i>Není třeba ho litovat; on si zaslouží</i>
<i>Jen tak dál</i>	<i>Je to dobré; Pokračujte prosím; Špatně to dopadne</i>
<i>Drž hubu</i>	<i>Uzavři se; buď zticha</i>

Zaprezentowana metoda synonimiczna, oparta na wyszukiwaniu wielowyrzowych jednostek odpowiadających znaczeniowo reproductowi wyjściowemu, okazała się przydatna w identyfikacji nowych fatyzmów i emotywów, umożliwiając tym samym poszerzenie zasobu wyrażen używanych w podobnych sytuacjach komunikacyjnych. Warto również zauważyć, że

metoda ta przyczynia się do odkrywania mniej oczywistych odpowiedników i wzbogaca badania nad językiem mówionym dzięki wykorzystaniu dostępnych internetowych słowników jednojęzycznych i synonimów. Procedura wyszukiwania jednostek, mimo swojej wyraźnej użyteczności, posiada pewne ograniczenia, wynikające głównie z nieprecyzyjnych wyników wyszukiwań spowodowanych ograniczeniami baz danych, takimi jak brak uwzględnienia wszystkich wariantów synonimicznych czy użycie dosłownych tłumaczeń, które bywają nieadekwatne w rzeczywistych kontekstach. Zróżnicowana frekwencja poświadczeń w korpusach językowych może prowadzić do błędnych wniosków dotyczących powszechności użycia wyrażen, szczególnie w przypadku jednostek potocznych, słabo reprezentowanych w tekstach pisanych. Ponadto, metoda ta wymaga każdorazowej weryfikacji odnalezionych odpowiedników, aby wykluczyć propozycje niezbieżne semantycznie z jednostką wyjściową. Istotne trudności napotkaliśmy w przypadku języka czeskiego, co jest związane z brakiem odpowiednich narzędzi umożliwiających analizę jednostek wielowyrazowych w internetowych słownikach synonimów. Oferowane przez wskazane platformy dane są niekompletne, ponieważ nie tylko proponują niedopasowane semantycznie lub stylistycznie odpowiedniki, ale przede wszystkim pomijają jednostki wielowyrazowe. Pomimo widocznych ograniczeń można uznać metodę synonimiczną za wartościową procedurę w badaniach nad jednostkami wielowyrazowymi, zwłaszcza w kontekście analizy języka mówionego (fatyzmów i emotywów). Niemniej jednak jej efektywne stosowanie wymaga każdorazowej odpowiedniej weryfikacji danych i uwzględnienia ograniczeń wynikających z różnorodności korpusów oraz narzędzi internetowych.

6.5 Wyszukiwanie przyrostowe

Przez wyszukiwanie przyrostowe, znane również jako *incremental search*, należy rozumieć metodę interaktywnego wyszukiwania informacji w systemach informatycznych, która umożliwia natychmiastową aktualizację wyników w odpowiedzi na kolejne wprowadzane znaki zapytania. Jest to technika powszechnie stosowana w aplikacjach takich jak edytory tekstu, przeglądarki internetowe czy wreszcie w wyszukiwarkach internetowych, które będą głównym obiektem naszego zainteresowania w ramach badań nad wielowyrazowymi jednostkami języka mówionego. W przeciwieństwie do tradycyjnego wyszukiwania, które wymaga pełnego sformułowania zapytania, wyszukiwanie przyrostowe umożliwia

użytkownikowi iteracyjne¹¹ precyzowanie zapytania w czasie rzeczywistym i bieżące otrzymywanie wyników w odpowiedzi na kolejno wpisywane przez użytkownika znaki. Mechanizm ten opiera się głównie na analizie wprowadzanych danych w celu przewidywania intencji użytkownika, co pozwala na szybsze dostarczanie adekwatnych wyników, które uwzględniają kontekst i personalizację otrzymanych rezultatów, np. po wpisaniu wyrażenia *nie chcę* wyszukiwarka internetowa sugeruje dalsze komponenty, które licznie rejestruje wskazana baza internetowa

→ *Nie chcę* → *Nie chcę Cię znać/Nie chcę, ale muszę* itp.

Kluczowym elementem są algorytmy autouzupełniania i odpowiedzi zapytań, które korzystają z analizy dużych zbiorów danych, takich jak popularne frazy czy historie wyszukiwań innych użytkowników. Jeden z ekspertów w dziedzinie interfejsów, amerykański projektant Jeff Raskin, w swojej pracy *The Humane Interface* w sposób następujący ocenił przydatność tego narzędzia: „Z punktu widzenia inżynierii interfejsów, zalety wyszukiwania przyrostowego są tak liczne, a zalety wyszukiwania ograniczonego tak nieliczne, że nie widzę prawie żadnych sytuacji, w których preferowane byłoby wyszukiwanie ograniczone” [tłum. własne] (Raskin 2000: 126).

Metoda wyszukiwania przyrostowego oferuje znacznie szerszy zakres zastosowań i możliwości niż wskazane powyżej, jednak w kontekście badań nad replektami języka mówionego, takimi jak fatyzmy i emotywy, szczególne znaczenie ma analiza danej funkcji w wyszukiwarkach internetowych, gdzie mechanizmy autouzupełniania i dynamicznego dostarczania wyników umożliwiają odnalezienie nowych jednostek wielowyrazowych bądź samych wariantów. Ponadto, jak wskazano powyżej, wyszukiwarki internetowe oferują najczęściej występujące odpowiedzi, skonstruowane na podstawie analizy zapytań użytkowników, co umożliwia wyekscerpowanie jednostek języka mówionego o potencjalnie wysokiej frekwencji. Owe narzędzia stanowią będą kluczowe instrumentarium w analizie każdego z wymienionych języków, pełniąc rolę podstawowego źródła do identyfikacji nowych jednostek wielowyrazowych. Przeprowadzona analiza będzie polegała na egzemplifikacyjnym określeniu najczęściej występujących elementów wcześniej zebranych fatyzmów i ematywów, wprowadzeniu ich do określonych wyszukiwarek, poszukiwaniu nowych jednostek oraz ustaleniu ich znaczenia semantycznego i częstości występowania w oparciu o zasoby korpusowe.

¹¹ W kontekście wyszukiwania przyrostowego iteracyjność oznacza, że użytkownik stopniowo precyzuje swoje zapytanie, modyfikując je w odpowiedzi na bieżąco wyświetlane wyniki. Dzięki temu proces jest dynamiczny i pozwala na lepsze dostosowanie wyszukiwania do oczekiwań i potrzeb użytkownika

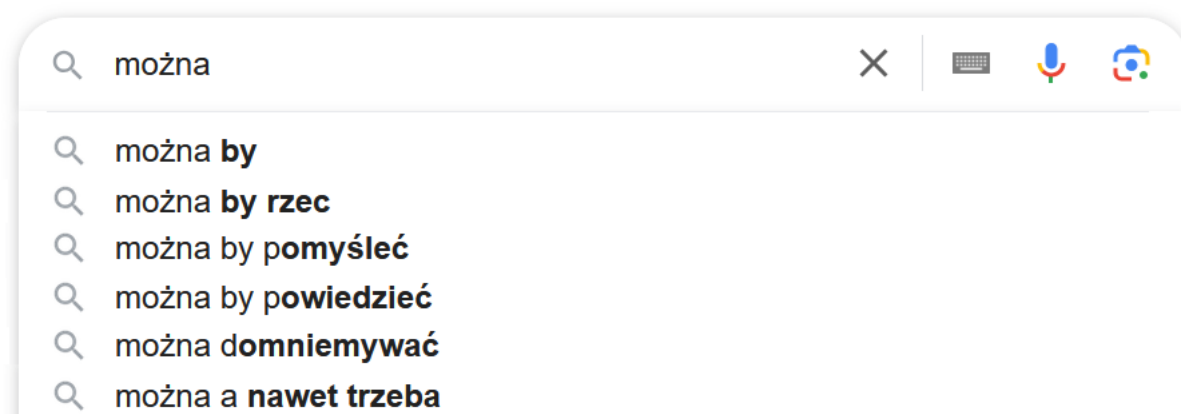
Wyszukiwarka *Google* w analizie fatyzmów i emotywów języka polskiego

W celu odnalezienia nowych przykładów polskich wielowyrazowych fatyzmów i emotywów zostanie wykorzystana funkcja wyszukiwania przyrostowego w domenie *google.com*. Według danych z 2024 roku opublikowanych przez StatCounter, firmę specjalizującą się w analizie statystyk internetowych, takich jak ruch w sieci, udział przeglądarek, wyszukiwarek i systemów operacyjnych, najpopularniejszą wyszukiwarką internetową na świecie pozostaje *Google*. Popularność Google w Polsce jest wyjątkowo wysoka – według różnych analiz jej udział w rynku wynosi ponad 95%, co wynika z prostoty obsługi, precyzji wyników czy integracji z innymi usługami ekosystemu Google.

Proces ten będzie przebiegał następująco:

1. Zidentyfikowanie częstych początkowych komponentów fatyzmów i emotywów;
2. Wpisanie danych komponentów do wyszukiwarki, a następnie kolejno dodawanie i usuwanie każdej litery alfabetu/bądź ich kombinacji, aby zaobserwować propozycje podsuwane przez wyszukiwarkę;
3. Wprowadzenie propozycji w odpowiednie korpusy oraz ustalenie ich frekwencji.

Dla celów ilustracyjnych wykorzystamy wyraz *można*, który charakteryzuje się wysoką frekwencją występowania jako komponent inicjalny w jednostkach fatycznych. Wyraz ten, ze względu na swoją uniwersalność i częste użycie w polskich strukturach językowych, stanowi właściwy przykład do obserwacji potencjalnych kombinacji jednostek wielowyrazowych, które będą oferowane przez narzędzie *Google*. Poniżej przedstawiono przykłady zaproponowanych przez wyszukiwarkę sugestii jednostek wielowyrazowych, które zostały zebrane i zaprezentowane w formie zbiorczej:



Założona analiza odnalezienia fatyzmów pozwoliła również uzyskać przykład komentarza reakcyjnego emotywnego, wyrażającego wstępną akceptację propozycji przedstawionej przez rozmówcę, co ilustruje pierwszy wynik wyszukiwania.

Wyrażenie *można by* znajduje wiele zastosowań w różnych kontekstach, takich jak wyrażanie hipotetycznej możliwości, wątpliwości czy ogólnej refleksji, a po nim najczęściej występują różnorodne, w zasadzie dowolne czasowniki, jak na przykład: *można by pomyśleć*; *można by spróbować* itp.

Wyszukiwarka udostępniła również parę wielowyrazowych jednostek synonimicznych *można by powiedzieć*/*można by rzecz*, które używane są przeważnie do wprowadzenia refleksji, opinii lub oceny mówiącego bądź to przedstawienia ogólnego sensu informacji w inny sposób. Druga z jednostek występuje częściej w tekstach literackich o charakterze stylizowanym i zdecydowanie rzadziej pojawia się w codziennej komunikacji.

Z kolei jednostka *można by pomyśleć* jest zwrotem, używanym głównie w celu wyrażenia hipotetycznego przypuszczenia lub opinii na temat jakiejś sytuacji, często pełniąc przy tym funkcję retoryczną, sugerując ironię bądź dystans wobec wyrażanego sądu. W niektórych kontekstach kolejny reprodkt *można domniemywać* może być stosowany synonimicznie w odniesieniu do poprzedniej jednostki, jednak jego podstawowe znaczenie ogranicza się do możliwości formułowania przypuszczeń na podstawie dostępnych przesłanek lub faktów, przy czym często zakłada pewien stopień niepewności i stanowi bardziej formalne wyrażenie. Ostatni z wyekscerpowanych zwrotów *można, a nawet trzeba* może pełnić funkcję zarówno fatyczną (wprowadzającą do wypowiedzi), jak i emotywną (reakcyjną i używaną w

celu podkreślenia, że określone działanie jest konieczne w określonych okolicznościach). Poniżej zaprezentowano przykłady użycia tych jednostek w wyodrębnionych ze zbiorów poświadczeniach tekstowych (NKJP, polska część InterCorp v16ud oraz Araneum Polonicum Maius) wraz z frekwencją użycia dla każdego z nich:

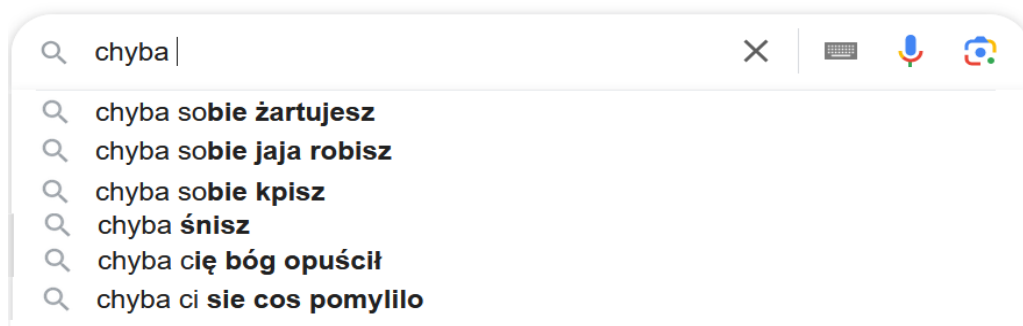
JEDNOSTKA	PRZYKŁAD W ZDANIU
<i>Można by</i>	<i>Nie była to zbyt przyjemna robota, ale na pytanie, dlaczego by nie użyć zwykłych nici lnianych, Rijen odparła: - Można by, ale wtedy skóra po brzegach się rwie i szwy będą przemakać.</i>
<i>Można by rzec</i>	<i>Jesteśmy u siebie, chociaż nie jest nam najlepiej. To trochę jak w „Białym” Krzysztofa Kieślowskiego, gdzie bohater prosto z zagranicy ląduje na śmietniku w walizce. Można by rzec, że żyjemy na śmietniku, ale w końcu to jest nasz śmietnik.</i>
<i>Można by pomyśleć</i>	<i>Bo czymże innym jest zbitka zdań: "Tych dwóch mężczyzn łączy coś jeszcze. Oczy. Z kryształkami lodu na dnie". Przecież nie chodzi tu po prostu o dwóch mężczyzn o lodowatym spojrzeniu. Można by pomyśleć, że te trzy zdania to przytoczona przez autorkę wypowiedź osoby opętanej przez szatana, ale tak też nie jest.</i>
<i>Można by powiedzieć</i>	<i>Odtamtąd jednak przestał się zupełnie pokazywać, w każdym razie nie tylko ja, lecz także inni nie widzieli go już więcej. Można by powiedzieć, że upragniony spokój nastał po nim.</i>
<i>Można domniemywać</i>	<i>Wystąpiła o zasilek, [...], choć nie podała na czym polegały jej kłopoty zdrowotne. Można domniemywać, że było to dość poważne schorzenie, ponieważ zdecydowała się wystąpić o rentę.</i>
<i>Można, a nawet trzeba</i>	<i>Czy konflikt można rozwiązać? Jak to zrobić? - Można a nawet trzeba. Posłużę się tutaj przykładem. W Gdańsku też jest konflikt. Głośny już w Polsce. Konflikt o kolor krzesełek na nowym stadionie.</i>

JEDNOSTKA	FREKWENCJA
<i>Można by</i>	138 726 poświadczeń
<i>Można by rzec</i>	3 171 poświadczeń
<i>Można by pomyśleć</i>	2 353 poświadczenia
<i>Można by powiedzieć</i>	6 466 poświadczeń
<i>Można domniemywać</i>	3 126 poświadczeń
<i>Można, a nawet trzeba</i>	158 poświadczeń

Uzyskane dane wskazują na wyraźne zróżnicowanie częstości użycia poszczególnych konstrukcji wyrażających możliwość, przy czym należy uwzględnić fakt, że najwyższa liczba poświadczeń dla jednostki *można by* odnosi się do jej znaczenia jako wyrażenia potencjalnej możliwości, najczęściej jednak w połączeniu z innymi wyrazami, a nie jako samodzielnego emotywu reakcyjnego. Wyniki potwierdziły zróżnicowane preferencje użytkowników w odniesieniu do synonimicznych wyrażen *można by powiedzieć* i *można by rzec*, przy czym drugie z nich występuje znacznie rzadziej w języku potocznym i jest odnotowywane w tekstach korpusowych prawie o połowę rzadziej. Mniej popularne wyekscerpowane frazy mogą wskazywać na ich bardziej specjalistyczny charakter lub ograniczoną funkcjonalność w określonych kontekstach, jednak zdecydowana większość z nich uwzględnia wyrażenie potencjalności czy hipotetyczności. Niska frekwencja wyrażenia *można, a nawet trzeba* wynika ponownie z ograniczonej reprezentacji tekstów mówionych w korpusach, podczas gdy wyrażenie to znacznie częściej występuje w języku ustnym niż w materiałach pisanych. Poniżej przedstawiono pozostałe przykłady, ilustrujące wpisanie początkowych elementów fatyzmów do wyszukiwarki *Google* i propozycje wyszukiwania przyrostowego:

JEDNOSTKA	GOOGLE.COM – WYSZUKIWANIE PRZYROSTOWE
<i>Tak czy</i>	<i>Tak czy siak; tak czy owak; tak czy inaczej</i>
<i>Ogólnie</i>	<i>Ogólnie mówiąc; Ogólnie rzecz biorąc; Ogólnie to</i>
<i>Jak by</i>	<i>Jak by nie było; Jak by na to nie patrzeć; Jak by nie patrzeć</i>

Po przeprowadzeniu próby identyfikacji jednostek fatycznych za pomocą wyszukiwania przyrostowego w domenie *google.com*, przechodzimy do analizy jednostek emotywnych. Obserwacja ich struktury w uprzednio zebranej bazie pozwoliła na wyciągnięcie wniosków, że częstym elementem występującym w takich jednostkach jest inicjalny wyraz *chyba*, pełniący w języku polskim głównie funkcję partykuły, wyrażającą niepewność lub wątpliwość mówiącego co do prawdziwości przekazywanej informacji. Jednak w kontekście wielowyrzowych jednostek emotywnych analizowany wyraz może zyskiwać zupełnie nowe znaczenie w wyniku połączenia z innymi jednostkami leksykalnymi i zanurzeniem w innym kontekście sytuacyjnym. Wpisanie wskazanego słowa do wyszukiwarki i naprzemienne dodawanie lub usuwanie kolejnych liter alfabetu (bądź kombinacji) przyniosło następujące rezultaty:



Narzędzie wyodrębniło ciąg wielowyrazowych jednostek synonimicznych w postaciach *chyba sobie żartujesz/chyba sobie jaja robisz/chyba sobie kpisz/chyba śnisz*, które są często używane w języku potocznym dla wyrażania zdziwienia lub dezaprobaty wobec wypowiedzi rozmówcy, której treść jest postrzegana jako niewiarygodna bądź absurdalna. Użycie danych fraz służy głównie zakwestionowaniu wiarygodności przekazu w sposób silnie nacechowany emocjonalnie i często skutkuje sprowokowaniem rozmówcy do ponownego przemyślenia swoich słów.

Podobną semantyką cechuje się wyrażenie *chyba Cię Bóg opuścił*, które również w sposób ekspresywny wyraża zdumienie lub krytykę wobec czyjejś wypowiedzi, szczególnie w kontekście zaproponowanego przez rozmówcę działania dotyczącego określonej kwestii. Fraza ta sugeruje, że adresat postępuje nieracjonalnie, co metaforycznie jest przedstawiane jako utrata boskiego wsparcia lub przewodnictwa.

Wysoką frekwencję w języku potocznym ma reprodukt emotywny *Chyba ci się coś pomyliło*, stanowiący reakcję podważającą trafność czyjejś opinii. Używany jest on zarówno w kontekście zarzucania rozmówcy błędnego przypisywania faktów, jak i w celu ironicznego wyrażenia krytyki wobec jego postawy. Przykłady użycia danych jednostek emotywnych, zaczerpnięte z tekstów dostępnych w internecie i zasobach korpusowych, zostały przedstawione poniżej wraz z informacją o ich częstotliwości występowania w poszczególnych konkordancjach:

JEDNOSTKA	PRZYKŁAD W ZDANIU
Chyba sobie żartujesz	<i>Ach, doprawdy, niewiele! Przynajmniej umorzenie tej drakońskiej podwyżki, czyli powrót do ceny pierwotnej, nominalnej. – Co? Do dwudziestu ośmiu?! – roześmiał się szyderczo. – Chyba sobie żartujesz!</i>
Chyba sobie jaja robisz	<i>- Czy wy w ogóle słuchaliście naszego demo? - Jakiego demo? Tego, które wysłałem do biura. Nie przesłali tego? - Nie. - Chyba sobie jaja robisz.</i>
Chyba sobie kpisz	<i>- Po prostu szybciej będzie. - Wolną ręką wyjąłem z worka plik dolarów, rzuciłem w lewo. Potem drugi, w prawo. - Twoja, moja, twoja, moja... Będzie was mniej, więc dostaniecie większe działki. To korzystny układ. - Chyba sobie kpisz - postawiła diagnozę Patrycja.</i>
Chyba śnisz	<i>Dziewczyno, zastanów się o czym Ty w ogóle mówisz. Mandaryna jest cool!? Chyba śnisz. Ona Zawsze „śpiewa” z playbacku, nigdy na żywo.</i>
Chyba Cię Bóg opuścił	<i>– Kiedy oddasz mi 5 dych? – Chyba Cię Bóg opuścił, oddałem dawno, sprawdź stan konta.</i>
Chyba Ci się coś pomyliło	<i>Ja nikogo nie obraziłam! I nie wywyższam się bo nawet nie czuję tego awansu, chyba Ci się coś pomyliło. I strasznie mnie rozśmieszyłaś teraz dopiero tym, że czuję się lepsza od innych.</i>

JEDNOSTKA	FREKWENCJA
<i>Chyba sobie żartujesz</i>	154 poświadczenia
<i>Chyba sobie jaja robisz</i>	57 poświadczeń
<i>Chyba sobie kpisz</i>	67 poświadczeń
<i>Chyba śnisz</i>	67 poświadczeń
<i>Chyba Cię Bóg opuścił</i>	3 poświadczenia
<i>Chyba Ci się coś pomyliło</i>	23 poświadczenia

Bardzo niska frekwencja występowania poszczególnych wyodrębnionych jednostek językowych wynika przede wszystkim z dwóch kluczowych czynników. Po pierwsze, analizowane reproduktory znajdują zastosowanie głównie w języku mówionym, który słabo jest dokumentowany w tekstach korpusowych. Po drugie, znacznie wyższa częstość występowania w materiałach tekstowych dotyczy skróconych wariantów analizowanych emotyów, takich jak przykładowo: *Żartujesz sobie?*; *Jaja sobie robisz?*, czy *Chyba kpisz!*; *Bóg Cię opuścił!*

W związku z tym konieczne jest uwzględnienie tych czynników zarówno w badaniach nad językiem mówionym, jak i w tworzeniu korpusów, które lepiej oddawałyby

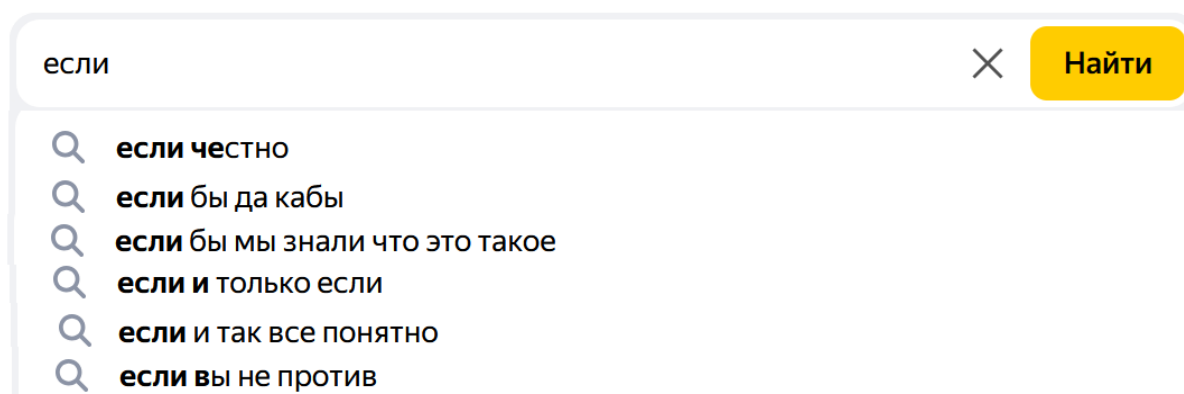
charakterystykę tego typu jednostek leksykalnych. Obecność danych wyrażen w mowie potocznej Polaków jest niezaprzeczalna, a ich ograniczona dokumentacja w korpusach może prowadzić do niedoszacowania ich znaczenia i nieuwzględnienia w opracowaniach leksykografii jedno- czy wielojęzycznej. Poniżej zaprezentowano dodatkowe przykłady ukazujące efekty wprowadzenia początkowych fragmentów emotywów do wyszukiwarki Google oraz związane z tym sugestie wyszukiwania przyrostowego:

JEDNOSTKA	GOOGLE.COM – WYSZUKIWANIE PRZYROSTOWE
Kto	<i>Kto to widział!; Kto by pomyślał!; Kto by się spodziewał!</i>
Nic	<i>Nic się wielkiego nie stało; Nic się nie stało; Nic nie szkodzi</i>
Takie	<i>Takie buty; Takie rzeczy; Takie sprawy</i>

Wyszukiwarka *Yandex* w analizie fatyzmów i emotywów języka rosyjskiego

Za dominującą wyszukiwarkę na rynku rosyjskim od lat uznaje się *Yandex.ru*, która, podobnie jak Google, oferuje szeroką gamę usług, takich jak mapy, poczta elektroniczna, tłumaczenia, a także dodatkowe narzędzia, np. platformy reklamowe, analityczne i chmurowe. Jej przewaga na tym rynku wynika z głębokiego zrozumienia specyfiki językowej i kulturowej Rosji, lepszego dostosowania algorytmów do lokalnych potrzeb użytkowników oraz długoletniej obecności w regionie, co pozwoliło na budowę silnej pozycji i lojalnej bazy użytkowników. Ponadto Yandex aktywnie rozwija nowe technologie, takie jak sztuczna inteligencja, co umacnia jej wizerunek jako lidera technologicznego na terytorium byłego ZSRR. Wyszukiwarka ta zostanie zatem naturalnie wykorzystana do próby identyfikacji fatyzmów i emotywów rosyjskich poprzez analogiczne wskazanie najczęściej występujących komponentów początkowych uprzednio zebranych jednostek oraz wprowadzenie ich do pola wyszukiwania.

Pierwsza z analiz zostanie poświęcona inicjalnemu elementowi fatycznemu *ес.и*, który w dosłownym użyciu służy do wprowadzania warunków lub hipotez w zdaniach, wyrażając potencjalne sytuacje i często łącząc się z innymi elementami gramatycznymi w celu budowania złożonych struktur warunkowych. Wpisanie danego słowa w wyszukiwarkę Yandex i naprzemienne uzupełnianie pola różnymi literami alfabetu rosyjskiego bądź jego kombinacjami przyniosło następujące rezultaty:



Pierwszy z zaproponowanych zwrotów *если честно* pełni przede wszystkim funkcję wtrącenia używanego do podkreślenia szczerości wypowiedzi bądź własnej opinii o osobistym charakterze, często w celu nawiązania bardziej bezpośredniego kontaktu mówiącego z rozmówcą.

Wyrażenie *если бы да кабы* odzwierciedla ironiczny lub humorystyczny stosunek mówiącego do hipotetycznych zdarzeń, podkreślając ich nierealność oraz często reprezentując postawę rezygnacji wobec określonej kwestii. Omawiany zwrot prawdopodobnie pojawił się w jednym z pierwszych wyników wyszukiwania ze względu na popularność filmu komediowego o tej samej nazwie z 2016 roku, przy czym zarówno tytuł filmu, jak i rezultat wyszukiwania stanowi skróconą wersję pełnego powiedzenia *если бы да кабы, да во рту росли грибы*.

Reprodukt *если бы мы знали, что это такое* charakteryzuje się bardziej złożoną etymologią, związaną z popularnością filmiku opublikowanego w 2016 roku przez kanał *Первый Тульский*. W nagraniu emerytka, wyrażając obawy przed tzw. śnieżnymi pchłami atakującymi wieś, wypowiedziała słowa: *Страшно, очень страшно. Мы не знаем, что это такое, если бы мы знали, что это такое*. Dane wyrażenie szybko zdobyło popularność, stając się w kulturze internetowej symbolem dezorientacji, niepewności oraz braku zrozumienia wobec nieznanых lub trudnych do wyjaśnienia zjawisk, a jego użycie zaczęło przybierać humorystyczny i ironiczny charakter.

Z kolei wielowyzrazowej jednostki *если и только если*, zaproponowanej przez wyszukiwarkę Yandex, nie można uznać za przykład interesującego nas fatyzmu, ponieważ konstrukcja ta występuje głównie w matematyce i wyraża konieczność spełnienia określonych kryteriów w celu zachowania określonych warunków logicznych.

Wielowyzrazowa jednostka *если и так всё понятно* jest przeważnie używana na końcu wypowiedzi w celu zasygnalizowania, że dodatkowe objaśnienia w pewnej kwestii nie są już

konieczne, ponieważ odbiorca prawdopodobnie zrozumiał wszystko na podstawie kontekstu lub wcześniejszych informacji. Często stosowana jest dla wyrażenia dezaprobaty, sugerując, że rozmówca już posiada pełnię wiedzy na dany temat, co czyni dalsze wyjaśnienia zbędnymi.

Jednostka *если вы не против* stanowi uprzejmą formułę stosowaną w celu uzyskania akceptacji lub zgody rozmówcy na podjęcie określonego działania bądź wtrącenia się do wypowiedzi i wyrażenia swojej opinii. Użycie w zdaniu każdej z zaproponowanych jednostek wraz z oszacowaną frekwencją ich użycia (NKJR, rosyjska część InterCorp v16ud i Araneum Russicum Maius) przedstawiono poniżej:

JEDNOSTKA	ПРИКЛАД В ЗДАНИУ
Если честно	<i>-Чего, любезный, веселитесь? -спросил капитан-шеф-повар. -Может, у вас план побега поспел?! -Если честно, то плана нету, - признался слоненок.</i>
Если бы да кабы	<i>— А вот на этот раз попался, — возразил Крячко. — И вообще ты так говоришь, потому что счастлив в супружестве. Еще неизвестно, как бы ты себя повел, если бы... — Если бы да кабы, — перебил его Гуров</i>
Если бы мы знали что это такое	<i>Мы не знаем что это такое, если бы мы знали что это такое, мы не знаем что это такое!</i>
Если и только если	<i>Классы, которые совпадают, — это один и тот же класс. Классы идентичны, если и только если их члены идентичны.</i>
Если и так всё понятно	<i>Я вот одного не пойму – зачем покупать курс Ваи, если и так всё понятно.</i>
Если вы не против	<i>— Вот и давай вместе ловить, — ответила бабушка Трошкина. — Не буду, это не мои мухи, — уперся дед. — Они из твоей комнаты налетели. — Если вы не против, я могла бы попробовать, — предложила Дарья.</i>

JEDNOSTKA	FREKWENCJA
<i>Если честно</i>	12 187 poświadczeń
<i>Если бы да кабы</i>	158 poświadczeń
<i>Если бы мы знали что это такое</i>	0 poświadczeń
<i>Если и только если</i>	37 poświadczeń
<i>Если и так всё понятно</i>	1 poświadczenie
<i>Если вы не против</i>	1023 poświadczenia

Przeprowadzona analiza potwierdziła wcześniejsze przypuszczenia, że wyrażenie *если и только если* nie może być zakwalifikowane do rejestru języka mówionego, ponieważ jest ono używane wyłącznie w kontekstach matematycznych, gdzie pełni funkcję precyzyjnego określenia relacji logicznych. Brak tego wyrażenia w zasobach korpusowych tekstów o charakterze codziennym wynika głównie z jego specjalistycznego charakteru oraz ograniczonego zakresu zastosowań poza dyskursem naukowym.

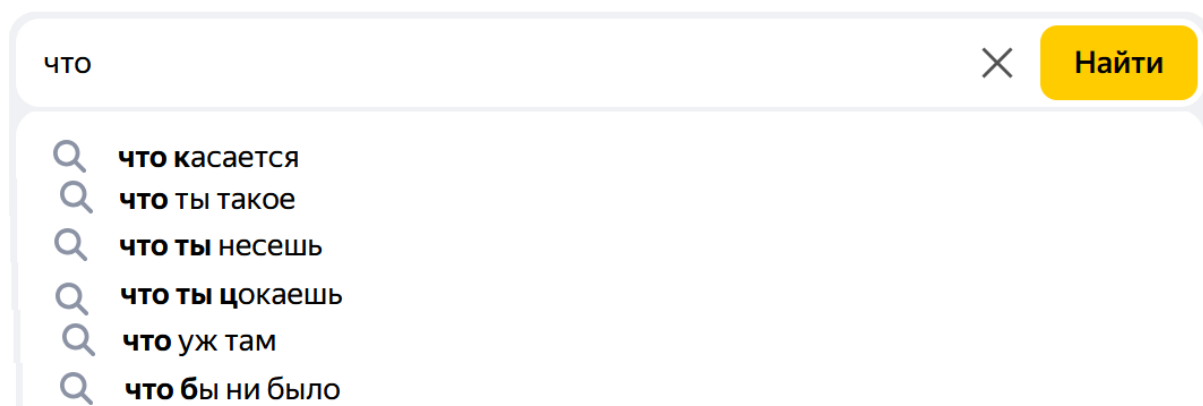
Nieobecność w korpusach językowych frazy *Если бы мы знали, что это такое* można tłumaczyć przede wszystkim znikomą reprezentacją tekstów internetowych w dostępnych zasobach, a także specyfiką rejestrowania skrzydlatych słów, które często występują w swoich niepełnych formach, co dodatkowo utrudnia ich rejestrowanie w zbiorach korpusowych.

Fraza *если и так всё понятно*, choć często pojawia się w wynikach wyszukiwań internetowych, jest praktycznie nieobecna w korpusach tekstowych, co wskazuje na różnicę między dynamicznym językiem internetowym a bardziej sformalizowanymi źródłami. Niezbędnym jest oszacowanie częstości występowania i kontekstów użycia w różnych odmianach języka rosyjskiego, ze szczególnym naciskiem na źródła języka internetowego. Poniżej przedstawiono przykłady innych wielowyrazowych jednostek języka mówionego, zaproponowanych przez daną wyszukiwarkę na podstawie wpisania początkowych, często występujących komponentów wcześniej zgromadzonych fatyzmów:

JEDNOSTKA	YANDEX.RU – WYSZUKIWANIE PRZYROSTOWE
<i>Вот</i>	<i>Вот так!; Вот это поворот; Вот так и живём</i>
<i>Почему</i>	<i>Почему бы и нет; Почему бы и да</i>
<i>Как</i>	<i>Как будто; как бы ни; как можно увидеть</i>

Ostatnia część analizy materiału w języku rosyjskim będzie polegać na próbie zidentyfikowania nowych wielowyrazowych jednostek emotywnych poprzez wyszukiwanie fraz rozpoczynających się od często występującego komponentu *Что*. Komponent ten, stanowiący element inicjalny w badanych jednostkach, pozwala na dołączanie do niego kolejnych elementów leksykalnych, tworząc różnorodne połączenia wyrazowe. Ze względu na swoje uniwersalne zastosowanie, *что* może występować w licznych kontekstach, co otwiera możliwość uzyskania szerokiego spektrum wyników, wymagających szczegółowej analizy w celu zidentyfikowania potencjalnych nowych jednostek językowych o charakterze

emotywnym. Wyekscerpowane przykłady jednostek wielowyrazowych, uzyskanych w wyniku wyszukiwania przyrostowego przedstawiają się następująco:



Pierwszy z zaproponowanych zwrotów stanowi przykład rozpowszechnionej jednostki fatycznej **что касается**, który jest stosowany jako komentarz wprowadzający do konkretnego tematu wypowiedzi lub wskazywania obiektu odniesienia przy sygnalizowaniu, że dalsza część wypowiedzi będzie dotyczyć określonej kwestii lub sytuacji.

Pojawienie się w wynikach wyszukiwania frazy **что ты такое** można wyjaśnić popularnością memu w rosyjskim internecie, powstałego na bazie sceny z filmu *Predator*, w której główny aktor, Arnold Schwarzenegger, wypowiada te słowa w reakcji na widok tytułowego Predatora. Wyrażenie to, wraz z memem, przekształciło się w skrzydlate wyrażenie funkcjonujące w przestrzeni internetowej, które jest wykorzystywane jako komentarz do obrazu lub informacji postrzeganych przez odbiorcę jako zupełnie niezrozumiałe, obce lub widziane po raz pierwszy.

Potoczne wyrażenie **что ты несёшь** jest używane dla wyrażenia zdziwienia lub niedowierzania wobec komunikatu rozmówcy, wskazując na jego absurdalność czy brak sensu, często pełniąc funkcję retoryczną bez oczekiwania odpowiedzi.

Z kolei reprodukt **что ты цокаешь** można wiązać z występem kabaretowym grupy *Уральские пельмени* w skeczu *Тинейджер*, w którym postać matki, reagując na niezadowolenie tytułowego nastolatka, wypowiada słowa: *Чё ты цокаешь? Чё ты глаза закатываешь? Чё ты вздыхаешь?*. Wyrażenie to, dzięki swojej ekspresji i humorystycznemu kontekstowi, zyskało popularność jako reakcja na okazywanie niezadowolenia przez innych, przenikając do języka potocznego jako ironiczny komentarz. Owa jednostka zazwyczaj jest rejestrowana w potocznym wariacie z użyciem słowa *чё*, a nie *что*.

Что уж там służy głównie wyrażeniu rezygnacji lub zaakceptowaniu czegoś, co jest już oczywiste lub nieuniknione, często w sytuacjach, gdy nadawca chce zminimalizować znaczenie tematu lub zakończyć dyskusję, wskazując na brak sensu dalszych rozważań. Ostatnie z wyrażen **что бы ни было** wskazuje na poddanie się jakimkolwiek okolicznościom przy wyrażeniu akceptacji i gotowości na każdą sytuację, która może się wydarzyć. Poniżej zaprezentowano przykłady użycia analizowanych jednostek w konkretnych poświadczeniach tekstowych wraz z ustaloną liczbą częstości występowania w zbiorach korpusowych:

JEDNOSTKA	ПРИКЛАД В ЗДАНИУ
Что касается	<i>Или ты считаешь, что если я покажу "статус" мне дадут автомашину или билет на поезд? Что касается опубликованного материала, он был опубликован только после того, как изучили этот вопрос со всех сторон.</i>
Что ты такое	<i>- Почему вы ко мне добры? Вопрос прозвучал жалобно и настороженно - боялся, что все его сочли слабаком-параноиком. Какое-то время Элен задумчиво молчала. - Я хочу понять ... кто ты. Что ты такое...</i>
Что ты несёшь	<i>– Мама, – вмешалась Джесси, – разве ты не видишь, что она расстроена? Для неё это то же самое, как для нас, если бы умер Черчилль. – Ну, что ты несёшь! – воскликнула шокированная тётушка и застыла на месте.</i>
Что ты цокаешь	<i>Чо ты цокаешь? Чо ты глаза закатываешь? Чо ты вздыхаешь?</i>
Что уж там	<i>После себя группа оставила бесчисленное количество альбомов – забудут их не скоро. Что уж там, расставаться всегда нелегко, особенно с этими ирландцами, не оставившими своих фанатов даже во времена Союза.</i>
Что бы ни было	<i>«Вы знаете, факт состоит в том, что нет богатых стран с высокой коррупцией . Поэтому что бы ни было , если Россия через 10 - 15 лет будет продолжать расти таким же темпом [...] коррупция исчезнет сама собой».</i>

JEDNOSTKA	FREKWENCJA
Что касается	88 248 poświadczeń
Что ты такое	103 poświadczenia
Что ты несёшь	391 poświadczeń
Что ты цокаешь	0 poświadczeń
Что уж там	560 poświadczeń
Что бы ни было	645 poświadczeń

Dane dotyczące liczby poświadczeń potwierdzają szerokie i uniwersalne zastosowanie wyrażenia *что касается* w kontekście jego funkcji jako elementu wprowadzającego do wypowiedzi oraz otwierającego nowy temat rozmowy. Wyrażenie to, będące charakterystycznym środkiem językowym, pełni istotną rolę w sytuacji oficjalnej i półoficjalnej, co znajduje odzwierciedlenie w jego częstej obecności w zasobach korpusowych.

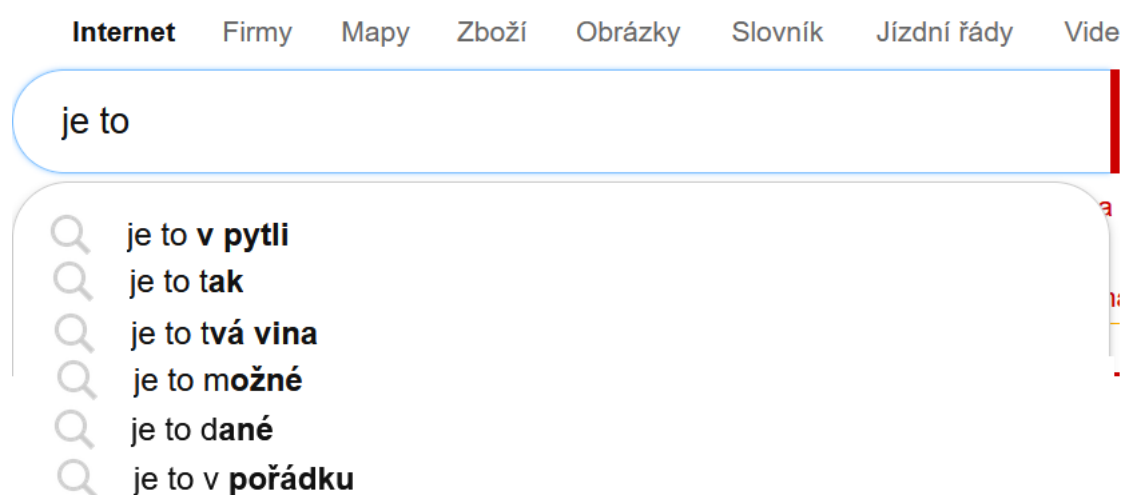
Stosunkowo niska frekwencja wyrażen potocznych, takich jak *что ты такое* oraz *что ты несёшь*, wynika przede wszystkim z ograniczonych zasobów języka mówionego w dostępnych korpusach językowych. Brak adekwatnego odzwierciedlenia tych konstrukcji w korpusach sprawia, że analiza ich rzeczywistego użycia w komunikacji codziennej staje się trudna. Podobną sytuację można zaobserwować w przypadku wyrażenia *Что ты цокаешь*, które pomimo relatywnie szerokiego zastosowania w dyskursie internetowym, nie jest odnotowywane w zasobach korpusowych. Ten fakt wskazuje na wyraźny brak w zasobach korpusowych kolokwializmów czy struktur charakterystycznych dla komunikacji codziennej i internetowej. Powyższe obserwacje ponownie unaoczniają potrzebę stworzenia lub znaczącego rozbudowania dedykowanego korpusu rejestrującego wyłącznie język potoczny, w tym także jego warianty stosowane w przestrzeni internetowej. Poniżej zaprezentowano przykłady pozostałych wielowyrazowych jednostek emotywnych, które udało się wyodrębnić przy pomocy funkcji przyrostowej w wyszukiwarce Yandex:

JEDNOSTKA	YANDEX.RU – WYSZUKIWANIE PRZYROSTOWE
<i>Это</i>	<i>Это всё; Это база; Это хорошая новость; Это вам не это</i>
<i>He</i>	<i>He твоё дело; He хочется верить; He может быть</i>
<i>Как</i>	<i>Как ты смеешь; Как ты вообще; Как это понять</i>

Wyszukiwarka Seznam w analizie fatyzmów i emotywów języka czeskiego

Za popularną czeską wyszukiwarkę uznaje się *Seznam.cz*, oferującą usługi takie jak mapy, poczta elektroniczna, wiadomości i ogłoszenia, które są ściśle dostosowane do lokalnych potrzeb i języka. Powstała w 1996 roku i początkowo funkcjonowała jako katalog stron internetowych. Z czasem Seznam rozwinął się w kompleksową platformę cyfrową, oferującą szeroki wachlarz usług dostosowanych do potrzeb czeskich użytkowników. Mimo silnej

konkurencji ze strony Google, Seznam utrzymuje dużą popularność dzięki dogłębnemu zrozumieniu specyfiki czeskiego rynku, integracji z lokalną kulturą oraz podobnie jak pozostałe wyszukiwarki jest również wyposażona w funkcję wyszukiwania przyrostowego. Należy jednak zauważyć, że wstępna analiza wykazała, iż proces poszukiwania wielowyrzowych jednostek języka mówionego w języku czeskim za pomocą wyszukiwarki Seznam jest bardziej utrudniony w porównaniu z wyszukiwarkami takimi jak Google czy Yandex. Seznam, po wpisaniu początkowych komponentów danej jednostki, kieruje podpowiedzi przede wszystkim na konkretne zapytania użytkowników obecne na forach internetowych, zamiast na potoczne jednostki wielowyrzowe. Zauważalne jest również sugerowanie wyszukiwań zawierających wskazany przez użytkownika wyraz w zapytaniach innych osób, co utrudnia łatwe wyodrębnienie interesujących nas jednostek. Próby wykorzystania tej wyszukiwarki do wyodrębnienia nowych wielowyrzowych jednostek fatycznych lub ich wariantów okazały się bezskuteczne, ponieważ wyszukiwarka w skrajnie rzadkich przypadkach proponuje wyłącznie pojedynczą sugestię stanowiącą przykład fatyzmu dla wskazanej jednostki, np. w przypadku słowa *pokud* wyszukiwarka zasugerowała jedynie wyrażenie *pokud možno*. Bardziej obiecujące wyniki przyniosła obserwacja i analiza jednostek emotywnych, które, choć również trudno wyodrębnić za pomocą wskazanej wyszukiwarki, okazują się możliwe do identyfikacji w przeciwieństwie do jednostek fatycznych. W celach egzemplifikacyjnych wykorzystane zostaną początkowe elementy ematywne o wysokiej frekwencji, takie jak *je to*, które posłużą jako dane wejściowe do wyszukiwarki w celu identyfikacji możliwych połączeń wyrazowych:



Wyekscerpowane wyrażenie *je to v pytlí* w swoim podstawowym znaczeniu odnosi się do sytuacji niekorzystnej lub przegranej, jednocześnie implikując brak możliwości rozwiązania danego problemu lub odwrócenia jego negatywnego przebiegu. Warto jednak zauważyć, że w

określonych kontekstach pragmatycznych fraza ta może przybierać znaczenie pozytywne, odnoszące się do pomyślnego zakończenia jakiejś sprawy lub osiągnięcia zamierzonego celu. Z kolei wieloznaczna jednostka **je to tak** pełni w wypowiedzi przede wszystkim funkcję potwierdzającą, wyrażając zgodność nadawcy z informacją przekazaną przez rozmówcę, ale może być również używana do wyrażenia rezygnacji lub pogodzenia się z sytuacją, w której nie ma już możliwości podjęcia skutecznych działań. Warto podkreślić, iż dane wyrażenie jest często wykorzystywane w wypowiedzi jako środek służący upewnieniu się nadawcy, czy poprawnie zrozumiał przekazaną mu treść.

Poprzez użycie wyodrębnionego emotywu **je to tvá vina** mówiący przypisuje rozmówcy odpowiedzialność za negatywny skutek określonej sytuacji lub problemu, wykorzystując konstrukcję o silnym nacechowaniu emocjonalnym, której struktura *je to <koho> vina* umożliwia przeniesienie odpowiedzialności na wskazaną osobę, jednocześnie modyfikując znaczenie wypowiedzi w zależności od kontekstu. Komentarz reakcyjny w postaci **je to možné** stanowi powszechnie stosowaną formę wyrażania potencjalnej możliwości zaistnienia określonego zdarzenia lub rozwiązania, sugerując, że jest ono realne w określonych warunkach, jednak bez wyrażania pewności ze strony mówiącego. Zaproponowana przez wyszukiwarkę jednostka **je to dané** praktycznie nie występuje w języku czeskim jako samodzielna replika, lecz najczęściej pojawia się w formie *je to dané tím, že...*, co pozwala zakwalifikować ją do jednostek fatycznych wprowadzających wypowiedź. W takim kontekście oznacza, że coś jest ustalone lub wynika z niezmiennych okoliczności, sugerując, że dany fakt nie podlega dyskusji i jest traktowany jako oczywisty. Ostatnie z wyrażen **je to v pořádku** wskazuje na akceptację lub zgodę, że nakreślona przez rozmówców sytuacja bądź działania z nią związane są właściwe i nie wymagają żadnych dodatkowych wyjaśnień czy zmian. Poniżej przedstawiono przykłady tekstowe ilustrujące funkcję wymienionych jednostek oraz łączną częstość ich występowania w zasobach tekstowych korpusu SYN v8, czeskiej części InterCorp v16ud oraz Araneum Bohemicum Maximum:

JEDNOSTKA	PRZYKŁAD W ZDANIU
Je to v pytli	<i>Minuli jsme zavřený obchod, potom veliké parkoviště kamionů, stanoviště přívěsů, dál byla jen tma tmoucí a potom osamělý dům. Už jsem si začínal myslet, že je to v pytli, když se to nakonec stalo.</i>
Je to tak	<i>Tak se mi zdá, mladý muži, že jste se nějak začal zajímat o mou Ettii. Je to tak, nebo se snad mýlím?" "Ano, je to tak, " odpověděl mladík. "V tom případě bych vám chtěl říct, že je to marné , poněvadž vás už někdo předešel.</i>

<i>Je to tvá vina</i>	<i>To ty jsi se mnou mluvil přes varhany! "pochopila náhle Valentýna. „Tys mi říkal, jak mám využít své schopnosti. To tvůj hlas jsem slyšela v hlavě. Je to tvá vina, že vypadám... takhle!</i>
<i>Je to možné</i>	<i>- Při výběru místa na nocování vždy dbejte na vlastní bezpečnost. - Mohu jet na dovolenou se svým psím kamarádem? Ano, je to možné. Prosíme vás však o provedení úklidu po vašem miláčkovi před vrácením vozu.</i>
<i>Je to dané</i>	<i>Je smutné, že nejplnější podniky jsou ty, kde se hraje muzika, která byla na výši za dob hlubokého totalismu. Ale je to dané tím, že dnešní generace ve většině případech nemá ani zdání, že lze poslouchat i jiný zvuk [...].</i>
<i>Je to v pořádku</i>	<i>Upřímně jsme řekli mamce, co děláme, abychom také nezemřeli nedostatkem slunečního svitu. Podívala se z Chrise na mě a slabě se usmála. „Je to v pořádku, ale ať se to nedozví babička. Ta by nesouhlasila, to dobře víte.</i>

JEDNOSTKA	FREKWENCJA
<i>Je to v pytlí</i>	739 poświadczeń
<i>Je to tak</i>	180 439 poświadczeń
<i>Je to tvá vina</i>	174 poświadczenia
<i>Je to možné</i>	68 515 poświadczeń
<i>Je to dané</i>	5 296 poświadczeń
<i>je to v pořádku</i>	23 166 poświadczeń

Na pierwszy rzut oka zauważalna jest znacząca różnica w uzyskanych poświadczeniach, wynikająca przede wszystkim z faktu, że niektóre ze wskazanych jednostek, jak już wspomniano, cechują się dużą wieloznacznością, a ich liczba poświadczeń nie odzwierciedla rzeczywistego użycia w znaczeniu komentarza emotywnego o charakterze reakcyjnym.

W przytoczonym spisie warte uwagi są przede wszystkim dwie jednostki:

1. *Je to tak* – w większości zaobserwowanych poświadczeń pełniła funkcję rozwijającą wypowiedź, po której następowało dowolne określenie zdaniowe i odwołujące komentarz do wcześniejszego kontekstu sytuacyjnego,

2. *Je to možné* – Najczęściej współwystępującymi komponentami z niniejszym wyrażeniem są elementy *pokud* i *jak*, które wpływają na zmianę jego znaczenia. W przypadku pierwszego z nich, *pokud je to možné* stanowi przykład wyrażenia wprowadzającego warunek i pełniącego funkcję komponentu fatycznego, natomiast w przypadku drugiego, *jak je to možné*

służy zapytaniu (wyrażeniu niedowierzania) dotyczącemu sposobu powstania określonej sytuacji.

Poniżej przedstawiono inne przykłady wielowyrazowych jednostek wyekscerpowanych za pomocą wyszukiwania przyrostowego, wykorzystującego często używane komponenty wcześniej zebranych jednostek emotywnych:

JEDNOSTKA	SEZNAM.CZ – WYSZUKIWANIE PRZYROSTOWE
<i>Snad</i>	<i>Snad se to uklidní; Snad to není možné; Snad to vyjde</i>
<i>Asi</i>	<i>Asi by to bylo; Asi by to šlo</i>
<i>Jak</i>	<i>Jak to vypadá?; Jak to je?; Jak to vidí</i>

Na podstawie przeprowadzonej analizy działania wyszukiwania przyrostowego w kontekście identyfikacji wielowyrazowych jednostek językowych, takich jak fatyzmy i emotywy, można sformułować kilka istotnych wniosków.

Wyniki wyraźnie ukazują zróżnicowaną skuteczność badanych wyszukiwarek w procesie ekscerpacji interesujących jednostek, co pozwala ocenić ich przydatność i efektywność w realizacji przyszłych badań nad fatyzmami i ematywami.

Wyszukiwarka Google okazała się skutecznym narzędziem w analizie wielowyrazowych jednostek języka polskiego. Dzięki zastosowaniu zaawansowanych algorytmów autouzupełniania oraz szerokiej bazy danych możliwa była identyfikacja nowych jednostek lub wariantów wcześniej zebranych. Jednak zauważalnym ograniczeniem podczas analizy jest brak pełnego wglądu w sposób funkcjonowania algorytmów podpowiedzi, co może utrudniać kompleksową obserwację sugerowanych wyników.

Podobnie skutecznym narzędziem, szczególnie w analizie języka rosyjskiego, była wyszukiwarka Yandex. Platforma ta, dzięki dobrej integracji z lokalnym kontekstem językowym i kulturowym, dostarczyła wartościowych danych na temat fatyzmów i ematywów używanych w mowie potocznej oraz w komunikacji internetowej. Jednak, analogicznie do Google, Yandex również nie zapewnia pełnej przejrzystości mechanizmów podpowiedzi, a dodatkowe ograniczenia związane ze znikomym zarejestrowaniem jednostek języka mówionego w dostępnych korpusach mogą prowadzić do nieuwzględniania rozpowszechnionych fraz w słownikach.

Najmniej efektywna okazała się wyszukiwarka *Seznam*, analizowana w odniesieniu do języka czeskiego. Jej funkcjonalność, skupiona na odzwierciedlaniu konkretnych zapytań użytkowników z forów internetowych, była niewystarczająca w identyfikacji

wielowyrzutowych jednostek fatycznych. W przypadku jednostek emotywnych wyniki były nieco lepsze, jednak wciąż znacznie ograniczone w porównaniu z rezultatami uzyskanymi przy użyciu Google i Yandex. Tym samym przydatność Seznam jako narzędzia badawczego w analizie czeskiego języka mówionego jest ograniczona. Z perspektywy badawczej wyszukiwarki internetowej, mimo wspomnianych ograniczeń, pozostają wartościowym narzędziem do wyodrębniania fatycznych i emotywnych form językowych, a automatyczne mechanizmy uzupełniania zapytania oferują możliwość identyfikacji nowych jednostek leksykalnych, które są często pomijane w tradycyjnych korpusach tekstowych. Istotne z punktu widzenia dalszej analizy wyekscerpowanych jednostek wydaje się opracowanie dedykowanych korpusów mowy potocznej, zwłaszcza dla języków słabiej reprezentowanych na globalnych platformach internetowych.

6.6 Metoda tłumaczeniowa

Jedną z najskuteczniejszych metod identyfikacji wielowyrzutowych jednostek języka docelowego jest metoda tłumaczeniowa, która opiera się na systematycznym przekładzie uprzednio wyselekcjonowanych jednostek językowych (w naszym przypadku fatyzmów i emotywów) na języki docelowe. Tego rodzaju podejście nie tylko umożliwia ustanowienie potencjalnych par przekładowych, ale także pozwala na znaczące poszerzenie bazy potocznych jednostek wielowyrzutowych w innych językach, uwzględniając ich specyfikę semantyczną, pragmatyczną i kulturową. Sposób ten umożliwia zarówno identyfikację różnic i podobieństw międzykulturowych i językowych, a także wnosi istotny wkład w badania nad wielojęzycznością, określając uniwersalne zasady tworzenia oraz rozumienia jednostek wielowyrzutowych. Ponadto, jej zastosowanie pozwala na odkrywanie nowych, dotychczas nieopisanych w leksykografii jednojęzycznej czy przekładowej analogicznych konstrukcji w określonych językach. Kluczowym elementem tego procesu jest zgromadzenie reprezentatywnej bazy emotywów i fatyzmów języka polskiego (wraz z ich definicjami) oraz ich wykorzystanie w procesie przekładu przy użyciu wybranych korpusów równoległych. Korpusy te nie tylko umożliwiają tłumaczenie jednostek, lecz także dostarczają szerokiego kontekstu treściowo-sytuacyjnego, pozwalającego na ocenę, czy dana jednostka mogłaby zostać użyta w analogicznej sytuacji komunikacyjnej w języku docelowym. W celu zilustrowania działania omawianej metody, wykorzystany zostanie wybrany fatyzm i emotyw języka polskiego, które następnie zostaną poddane analizie w wybranych korpusach równoległych pod kątem ich przekładu na język rosyjski i czeski. Przeprowadzone badanie

umożliwi wyodrębnienie tłumaczeń proponowanych przez bazy tekstowe oraz wyekscerpowanie jednostek, które potencjalnie mogą pełnić analogiczną funkcję komunikacyjną w zbliżonych sytuacjach i kontekstach wypowiedzi.

Pierwszą z analizowanych jednostek będzie wprowadzający do wypowiedzi komunikat *nawiasem mówiąc*, który pełni głównie funkcję nawiązania lub podtrzymania kontaktu poprzez zasygnalizowanie wtrącenia lub wprowadzenia dygresji do rozmowy. Do analizy polsko-rosyjskiej został wykorzystany polsko-rosyjski korpus równoległy dostępny w NKJR oraz korpus InterCorp v16ud. Należy podkreślić, iż zasadniczym przedmiotem zainteresowania były wielowyrazowe ekwiwalenty jednostek, nawet w przypadkach, gdy ich najczęstszym odpowiednikiem w języku docelowym była forma jednowyrazowa. Dla przedstawionych poniżej przykładów odpowiedników oszacowano stopień frekwencji w oparciu o zbiory NKJR, rosyjskiej części InterCorp v16ud oraz korpusu Aranenum Russicum Maius. Analiza wybranej jednostki fatycznej we wskazanych bazach tekstowych dostarczyła następujących wyników, uwzględniających wyekscerpowany kontekst jej użycia, zaproponowane tłumaczenia oraz łączną liczbę poświadczeń:

Nawiasem mówiąc – Между прочим	
- Jakiś czas temu BBC przygotowała historyczny materiał na temat Churchilla, nawiasem mówiąc , zagorzałego katolika.	- Так вот. Би-би-си давала биографический материал об этом человеке. Черчилль, между прочим , был глубоко верующим католиком.

Nawiasem mówiąc – К слову сказать	
Ernazar, nawiasem mówiąc , lubił pogawędzić, zadawać pytania. Był postawny, służył w oddziałach kawaleryjskich, ostatnich po wojnie.	Эрназар, к слову сказать , любил порассуждать, порасспросить. И собой был видный малый, в армии, оказывается, служил в последних кавалерийских частях после войны.

Nawiasem mówiąc – Кстати сказать	
Nawiasem mówiąc , był to człowiek nader postawny, wysoki, trochę łysy, trochę szpakowaty i dosyć tęgi [...]	Кстати сказать : человек он был собою видный, осанистый, росту высокого, немного лыс [...]

JEDNOSTKA	FREKWENCJA
<i>Между прочим</i>	43 157 poświadczeń
<i>К слову сказать</i>	3 397 poświadczeń
<i>Кстати сказать</i>	4 238 poświadczeń

Na podstawie analizy danych przedstawionych w tabelach można stwierdzić, iż między wyekscerpowanymi ekwiwalentami *между прочим*, *к слову сказать* i *кстати сказать*, a wybraną jednostką polską zachodzi zgodność semantyczna i ich zastosowanie w tożsamyh sytuacjach komunikacyjnych.

W przypadku pierwszej jednostki często obserwuje się jej tłumaczenie na język polski jako *między innymi*, które w wielu kontekstach jest zastosowane prawidłowo, jednak w omawianym badaniu zachowana została funkcja wtrącenia do wypowiedzi rozmówcy, nie zmieniając przy tym głównego toku rozmowy. Stosunkowo niska frekwencja pozostałych jednostek wynika głównie z faktu, iż najczęstszym tłumaczeniem rosyjskim dla reprodktu *nawiasem mówiąc* jest odpowiednik jednowyrazowy *кстати*, który, jak można zauważyć, stanowi początkowy komponent trzeciej z wyodrębnionych w korpusach jednostek. W taki sposób przy użyciu jednej jednostki fatycznej polskiej uzyskaliśmy trzy dodatkowe rosyjskie jednostki wielowyrazowe, które mogą być ekwiwalentami wskazanej jednostki w odpowiednich sytuacjach komunikacyjnych, w zależności od kontekstu i nacechowania stylistycznego wypowiedzi. Przeprowadźmy teraz analogiczną analizę dla języka czeskiego, wykorzystując korpus równoległy InterCorp v16ud do identyfikacji czeskich ekwiwalentów oraz korpusy Synv8, czeską część InterCorp i korpus Araneum Bohemicum Maximum do obliczenia frekwencji dla wyekscerpowanych jednostek. Wyniki przeprowadzonej analizy, które uwzględniają zarówno identyfikację tłumaczeń, jak i analizę częstości ich użycia w poszczególnych bazach tekstowych, zostały zaprezentowane poniżej:

Nawiasem mówiąc – Dodejme, že...	
<i>Nawiasem mówiąc, odnalazł się z powrotem w swoim garniturze natychmiast po wkroczeniu milicji do jego gabinetu, ku nieprzytomnej radości jego sekretarki i ku nieopisanemu zdumieniu niepotrzebnie wezwanej milicji.</i>	<i>Dodejme, že k nepřičetné radosti Anny Richardovny a k velkému údivu zbytečně přivolaných orgánů vklouzl hbitě do svého proužkovaného obleku, sotva milice stanula ve dveřích.</i>

Nawiasem m $\acute{o$ wiac – Kdy $\acute{z}$ o tom mluv $\acute{i}$ m	
<i>Nawiasem m$\acute{o$wiac</i> ogromnie lubilem patrzeć na Kulikowa we wszystkich takich razach, to znaczy we wszystkich wypadkach, kiedy si $\acute{e}$ musiał postawić.	<i>Kdy$\acute{z}$ o tom mluv$\acute{i}$m:</i> velmi rád jsem se na Kulikova díval v podobných případech, totiž ve všech třech případech, kdy se musel ukázat.

Nawiasem m $\acute{o$ wiac – Jen tak mimochodem	
<i>Nawiasem m$\acute{o$wiac:</i> obchodzę dziś trzydzieste urodziny.	<i>Jen tak mimochodem:</i> absolvuji dnes své třicáté narozeniny.

JEDNOSTKA	FREKWENCJA
<i>Dodejme, že</i>	355 poświadczeń
<i>Kdy$\acute{z}$ o tom mluv$\acute{i}$m</i>	14 177 poświadczeń
<i>Jen tak mimochodem</i>	12 107 poświadczeń

Nie ulega wątpliwości fakt, iż wyekscerpowane z poświadczeń tekstowych czeskie tłumaczenia pod względem swojej semantyki i zastosowania w wypowiedzi pełnią silnie zbliżoną funkcję wtrącenia i uzupełnienia wypowiedzi, podobnie do *nawiasem m $\acute{o$ wiac*. Najczęściej odnotowanymi ekwiwalentami wskazanej jednostki wyjściowej okazały się odpowiedniki jednowyrazowe, takie jak *ostatně* albo *mimochodem*, wśród których drugi stanowi integralną część trzeciego z wyodrębnionych czeskich reproduktów fatycznych, tj. *jen tak mimochodem*. Z kolei jednostka *dodejme, že* odznacza się stosunkowo niską liczbą poświadczeń, co można tłumaczyć jej częstszym występowaniem w sytuacji oficjalnej oraz rzadkim użyciem w języku codziennym, a także mniejszą uniwersalnością w porównaniu z pozostałymi jednostkami. Dodatkowo, preferencje użytkowników języka czeskiego do stosowania bardziej rozpowszechnionych form synonimicznych mogą wpływać na ograniczoną frekwencję tego wyrażenia. Niemniej jednak przeprowadzona analiza pozwoliła na identyfikację trzech dodatkowych wielowyrazowych jednostek fatycznych w języku czeskim, a jednocześnie dostarczyła wiarygodnych przekładów dla wskazanej jednostki wyjściowej.

Przejdźmy zatem do próby wyodrębnienia nowych jednostek emotywnych w językach rosyjskim i czeskim, wykorzystując w tym celu polską jednostkę *nie mam pojęcia*, która służy do wyrażenia braku wiedzy oraz całkowitej nieświadomości mówiącego w odniesieniu do pytania zadanego przez rozmówcę. Zastosowanie niniejszej jednostki do identyfikacji

emotywów rosyjskich w tych samych źródłach, co w poprzednim etapie analizy, pozwoliło uzyskać następujące ekwiwalenty przekładowe wraz z odpowiadającą im liczbą poświadczeń:

Nie mam pojęcia – Понятия не имею	
<i>To jak sądzisz, dlaczego twój ojciec został zamordowany? Poczula, jak napinają się jej mięśnie. - Nie mam pojęcia.</i>	<i>- В таком случае ... почему ты полагаешь, что твой отец убит? - Понятия не имею, - сказала девушка, напрягшись всем телом.</i>

Nie mam pojęcia – Даже не представляю	
<i>- Proszę księdza! Nie ma go. Spakował swoje rzeczy i odszedł. Dokąd mógłby pójść? - Nie mam pojęcia.</i>	<i>- Отец! Его нет! Он собрал все свои вещи. Он ушел. Куда же он пошел? - Даже не представляю.</i>

Nie mam pojęcia – Без понятия	
<i>Jak myślisz, czego on szuka?! - krzyknął Ron. - Nie mam pojęcia! - odkrzyknął Harry. Ręce zdreptały mu z zimna, ale nie śmiał zwolnić uścisku.</i>	<i>— Как вы думаете, чего он ищет? — прокричал Рон. — Без понятия! — крикнул Гарри в ответ. Руки у него заоченели от холода, но он не решался ослабить хватку.</i>

Nie mam pojęcia – Не знаю	
<i>- Możliwe, że w Brazylii. - A dokładnie gdzie? Stephano zdobył się na lekceważące wzruszenie ramionami. - Nie mam pojęcia. To wielki kraj.</i>	<i>— В Бразилии, вероятно. — Где именно в Бразилии? Стефано вяло пожал плечами: — Не знаю. Бразилия — большая страна.</i>

Nie mam pojęcia – Представления не имею	
<i>-Po co, do wszystkich diabłów, kupił to gówno? - Nie mam pojęcia. Szacunkowo sprzedaż tych obiektów może dać coś ponad cztery miliony koron.</i>	<i>- Зачем, скажи на милость, он обзавелся этим дерьмом? - Представления не имею. По предварительным оценкам, наследство может при продаже составить четыре с лишним миллиона.</i>

JEDNOSTKA	FREKWENCJA
<i>Не знаю</i>	291 955 poświadczeń
<i>Не имею понятия</i>	388 poświadczeń

<i>Даже не представляю</i>	727 poświadczeń
<i>Без понятия</i>	2 417 poświadczeń
<i>Представления не имею</i>	115 poświadczeń

Zasoby NKJR i korpusu polsko-rosyjskiego InterCorp v16ud dostarczyły aż pięć wielowyzrazowych ekwiwalentów przekładowych, wśród których najwyższą frekwencją użycia cechuje się najbardziej uniwersalne stwierdzenie *не знаю*, używane dzięki swojej neutralności stylistycznej w każdej sytuacji. Nie ulega wątpliwości fakt, że wyekscerpowane przykłady odpowiedników tłumaczeniowych stanowią wierne ekwiwalenty jednostki wyjściowej *nie mam pojęcia*. Należy jednak podkreślić, że pewne zwroty wykazują drobne różnice. Przykładowo, *понятия не имею* jest zwrotem stosowanym głównie w języku potocznym, który znaczeniowo jest bardzo bliski jednostce wyjściowej, często jednak obarczonym zabarwieniem emocjonalnym. Z kolei *даже не представляю* oraz *представления не имею* są bardziej charakterystyczne dla sytuacji formalnych i oficjalnych, co czyni je częściej spotykanymi w tłumaczeniach dostarczonych przez korpusy w kontekstach wymagających większej precyzji lub dystansu.

Без понятия to zwrot potoczny, wręcz kolokwialny, najczęściej rejestrowany w przekładach z korpusu OpenSubtitles, gdzie użytkownicy wykazują preferencję do stosowania potocyzmów, odpowiadających specyfice sytuacji komunikacyjnej w dialogach filmowych czy serialowych. Przejdźmy zatem od analizy polsko-rosyjskiej do prób wyekscerpowania reproduktów emotywnych języka czeskiego oraz wskazania dla nich odpowiedników tłumaczeniowych oraz liczby poświadczeń, opierając się na wcześniej wskazanych czeskich źródłach tekstowych.

Nie mam pojęcia – Nemám tušení	
<i>Ba! Wiesz – zapytał gorączkowo – co zrobiłem potem? – Nie mam pojęcia, Ford. – Szkoda – rzekł Ford – miałem nadzieję, że mi przypominasz.</i>	<i>Tak! Co myslíš, že jsem udělal pak? "Nemám tušení. " Škoda, " zalitoval Ford, " doufal jsem , že mi to připomeneš.</i>

Nie mam pojęcia – Nemám poněť	
<i>Zadzwonil na jeden z prywatnych wewnętrznych numerów watykańskich. Nie mam pojęcia, skąd</i>	<i>Volá na jednu z vatikánských soukromých linek. Nemám poněť, kde to číslo vzal. " Rocher se</i>

go wziął. Rocher zatrzymał się. - Co takiego ? - Chce rozmawiać wyłącznie z dowódcą.	zastavil. "A co říká? " Chce mluvit jenom s velicím důstojníkem.

Nie mam pojęcia – Nemám zdání	
Pamiętał, że pracował w tej sali przez większą część feralnego dnia. Randolph pochylił się i spytał przyciszonym głosem: - Do czego ona zmierza? - Nie mam pojęcia – wyszeptał.	Vzpomněl si, že pracoval na sále číslo jedenáct skoro celý den. Randolph se k němu naklonil a zeptal se tlumeně: "Kam míří? " Nemám zdání , " zašeptal Jeffrey.

Nie mam pojęcia – Nemám páru	
- Tak, Taylor Swift uratowała przemysł muzyczny! - Kto to cholera jest? - Nie mam pojęcia.	- Taylor Swift zachránila hudební průmysl! - Kdo to k čertu byl? - Nemám páru.

JEDNOSTKA	FREKWENCJA
<i>Nemám tušení</i>	8 611 poświadczeń
<i>Nemám poněti</i>	6 453 poświadczenia
<i>Nemám zdání</i>	1 171 poświadczeń
<i>Nemám páru</i>	1 183 poświadczenia

W wyniku analizy wyodrębniono najczęściej używaną jednostkę, używaną zarówno w sytuacjach oficjalnych, jak i nieoficjalnych, którą jest ***nemám tušení***, stanowiącą wierny odpowiednik przekładowy zadanej jednostki wyjściowej. Jednostką o tożsamej semantyce i stosunkowo wysokiej liczbie poświadczeń, zawierającą dodatkowo odcień zdziwienia lub wątpliwości, jest kolejny wyekscerpowany reprodukt ***nemám poněti***, który w tym przypadku stanowi dosłowny przekład jednostki *nie mam pojęcia*. Niska liczba poświadczeń uzyskanego emotywu ***nemám zdání*** wynika najprawdopodobniej z jego rzadkiego użycia w sytuacjach codziennych, co można przypisać jego bardziej starannemu charakterowi.

Ostatni z reproduktów ***nemám páru***, uzyskany z bazy OpenSubtitles, dominuje w języku potocznym i jest często używany z nutą ironii lub humoru, dlatego można przypuszczać, że w języku polskim powinien być stosowany w analogicznym rejestrze.

Na podstawie przeprowadzonej analizy można stwierdzić, że zaprezentowana metoda znacząco wspiera badania nad reproduktami, wielojęzycznością oraz identyfikacją

ekwiwalentów wielowyrazowych jednostek językowych w różnych językach. Jej największą zaletą jest precyzyjne zestawianie jednostek wyjściowych z potencjalnymi odpowiednikami w językach docelowych. Proces tłumaczeniowy, przeprowadzany z wykorzystaniem korpusów równoległych, umożliwia uwzględnienie zarówno semantycznych, jak i pragmatycznych aspektów użycia analizowanych wyrażen w różnych kontekstach. Kluczowe zadanie, jakim jest odkrywanie nowych jednostek leksykalnych, które mogłyby pozostać niezarejestrowane w tradycyjnych opracowaniach leksykograficznych, zarówno jedno-, jak i wielojęzycznych, zostało zrealizowane na przedstawionych przykładach. Zaletą metody jest również analiza międzykulturowych podobieństw i różnic oraz identyfikacja jednostek o różnym nacechowaniu stylistycznym, co wspiera rozwój wiedzy o uniwersalnych mechanizmach językowych i znajduje zastosowanie w tłumaczeniach jednostek potocznych. Niemniej jednak, głównym ograniczeniem wskazanej metody jest zależność od dostępnych zbiorów korpusów równoległych, które często nie rejestrują przekładów jednostek potocznych lub ograniczają się do najpopularniejszych, zwykle jednowyrazowych odpowiedników, pomijając inne potencjalne ekwiwalenty. Na koniec warto zauważyć, że omawianą metodę można zintegrować z podejściem tłumaczenia dwukierunkowego, polegającym na ponownym wprowadzeniu wyekscerpowanych jednostek rosyjskich i czeskich do korpusu w celu ich tłumaczenia na język polski, co nie tylko umożliwia rozbudowanie bazy polskich jednostek, ale również pozwala na weryfikację poprawności uzyskanych par przekładowych. Wyekscerpowane w ramach przeprowadzonej analizy jednostki rosyjskie i czeskie zostały wprowadzone do korpusu InterCorp, co pozwoliło na uzyskanie następujących polskich jednostek emotywnych i fatycznych:

JEDNOSTKA ROSYJSKA	POLSKIE EKWIWALENTY
<i>Между прочим</i>	<i>Mimochodem; Nawiasem mówiąc; Między innymi</i>
<i>К слову сказать</i>	<i>Skoro już o tym mowa; A tak w ogóle; Z drugiej strony</i>
<i>Кстати сказать</i>	<i>No i jeszcze jedno; Zresztą; Przy okazji</i>
<i>Не знаю</i>	<i>Nie wiem; Naprawdę nie wiem</i>
<i>Не имею понятия</i>	<i>Nie mam pojęcia; Nie mam zielonego pojęcia</i>
<i>Даже не представляю</i>	<i>Nawet nie wiem; Nie wyobrażam sobie</i>

<i>Без понятия</i>	<i>Mnie nie pytaj; A bo ja wiem; Zupełnie nie wiem</i>
<i>Представления не имею</i>	<i>Nie mam pojęcia</i>

JEDNOSTKA CZESKA	POLSKIE EKWIWALENTY
<i>Dodejme, že</i>	<i>Dodać trzeba, že; Należy tu dodać; Nawiasem mówiąc</i>
<i>Když o tom mluvím</i>	<i>Nawiasowa uwaga; Nawiasem mówiąc; Jeśli o tym mowa</i>
<i>Jen tak mimochodem</i>	<i>A tak na marginesie; A przy okazji; Swoją drogą</i>
<i>Nemám tušení</i>	<i>Nie wiem; Nie mam pojęcia; Nie wiem, o czym mówisz</i>
<i>Nemám poněti</i>	<i>Nie mam żadnego pojęcia; Pojęcia nie mam</i>
<i>Nemám zdání</i>	<i>Nie mam pojęcia; Rany, nie wiem;</i>
<i>Nemám páru</i>	<i>Nie mam zielonego pojęcia; Nie czaję</i>

ROZDZIAŁ VII

Metody ustalania ekwiwalentów tłumaczeniowych dla jednostek fatycznych i emotywnych

Reprodukty fatyczne i emotywne, jako formuły mocno osadzone w języku mówionym, stanowią jedno z największych wyzwań w procesie tłumaczenia. W niniejszym rozdziale omówione zostaną wybrane metody ustalania ich ekwiwalentów przekładowych w relacjach polsko-rosyjskich i polsko-czeskich – zarówno korpusowe, jak i oparte na analizie materiałów równoległych, audiowizualnych oraz słownikowych.

Punktem wyjścia jest zatem rozumienie przekładu jako działania tekstowego, obejmującego zarówno sam proces, jak i jego rezultat (Lewicki 2022: 15). Jest to operacja odnosząca się do całych wypowiedzi, w których ustala się znaczeniowo-komunikacyjne zależności między jednostkami należącymi do różnych języków.

Odtwarzalność, obok trwałości strukturalnej i integralności semantycznej, stanowi podstawę definicji frazemu (Шанский 1957: 116), uwzględniającej także idiomy, skrzydlate słowa czy formuły etykietalne. Wszystkie te wyrażenia cechuje również sytuacyjność – są bowiem powtarzane w określonych warunkach komunikacyjnych, realizując konkretne intencje mówiącego. Tłumaczenie reproduktów stanowi zatem poważne wyzwanie, ponieważ nawet ustalone ekwiwalenty mogą znacząco różnić się pod względem składu komponentów. W praktyce rzadko udaje się uzyskać przekład, który zachowuje ekwiwalencję we wszystkich istotnych aspektach: obrazowym, dosłownym, emocjonalnym, stylistycznym oraz narodowo-etnicznym (Charciarek 2016: 743). Z racji uznania odtwarzalności i powtarzalności za cechy konstytutywne reproduktów, dobór ekwiwalentu powinien uwzględniać również analizę ich częstości i funkcjonowania w rzeczywistej komunikacji. Służy temu między innymi analiza danych korpusowych, która umożliwia empiryczną ocenę użycia reproduktów oraz identyfikację ich funkcjonalnych odpowiedników. Jak pisze František Čermák: „Empirismus je obecně vědecký přístup vycházející z objektivních dat, obvykle masových, a informace o nich, který stojí v protikladu k racionalismu, který se naproti tomu opírá o osobní, a tedy i nutně subjektivní introspekci” (Čermák 2017: 27).

Jednym z centralnych pojęć w teoriach translatorycznych, zwłaszcza tych o orientacji językoznawczej, jest ekwiwalencja. W ujęciu ogólnym termin ten można określić jako relację

znaczeniowo-funkcjonalną między jednostką językową w języku wyjściowym (L1) a odpowiadającą jej jednostką w języku docelowym (L2).

Należy podkreślić, iż ekwiwalencja przekładowa nie polega na prostym przeniesieniu znaczeń, lecz na ich świadomym doborze, tj. tłumacz wybiera znaczenie zbliżone do oryginału, choć nie zawsze z nim tożsame (Lewicki 2022: 145–146).

Jedną z klasycznych typologii ekwiwalencji zaproponował Eugene Nida, który wyróżnił ekwiwalencję formalną oraz dynamiczną. Pierwsza z nich zakłada maksymalne zachowanie struktury językowej oryginału (zarówno pod względem formy, jak i znaczenia) i wiąże się najczęściej z tłumaczeniem dosłownym. Z kolei druga opiera się na funkcjonalnym dopasowaniu jednostki przekładu do oryginału tak, aby wywołać w odbiorcy języka docelowego podobny efekt jak w odbiorcy tekstu źródłowego. W ramach tego podejścia wyróżnia się również m.in. ekwiwalencję funkcjonalną, polegającą na zachowaniu funkcji tekstu oraz ekwiwalencję pragmatyczną, w której tłumaczenie osiąga zamierzony przez autora efekt komunikacyjny. To właśnie ekwiwalencja pragmatyczna, rozumiana jako zgodność intencji, funkcji i oddziaływania, stanowi kluczowy punkt odniesienia w przekładzie jednostek fatycznych i emotywnych, gdzie zachowanie efektu komunikacyjnego bywa ważniejsze niż zgodność strukturalna (Jedlička 2019: 16).

W literaturze przedmiotu spotkać można jeszcze inne typologie stopni ekwiwalencji frazeologicznej.

Rosemarie Gläser wyróżnia ekwiwalencję pełną, częściową, brak ekwiwalencji oraz tzw. fałszywych przyjaciół tłumacza (Blagoeva, Sosnowski 2022: 51), natomiast Mona Baker klasyfikuje idiomy według ich obecności w języku docelowym, zakresu znaczeń i zgodności kontekstowej (Baker 1992: 65–71).

Na szczególną uwagę zasługuje propozycja Ludmiły Stěpanovej, która wyróżnia pięć typów ekwiwalentów:

- pełne (identyczne pod względem struktury, obrazowości i znaczenia),
- częściowe (zbieżne semantycznie i obrazowo, lecz różne formalnie),
- względne (z modyfikowaną obrazowością),
- analogie frazeologiczne (inne strukturalnie, lecz zbliżone znaczeniowo i stylistycznie) oraz
- jednostki bezekwiwalentne, silnie powiązane z kulturą (Stěpanova 2014: 134).

Istotny wkład w rozwój teorii przekładu wniosła również koncepcja K. Reissa i H. Vermeera, według której kluczowe jest osiągnięcie celu komunikacyjnego, a ekwiwalencja

stanowi jedynie jego szczególny przypadek (Nevzorova-Kmech 2022: 51). O sukcesie przekładu decyduje zatem adekwatność, czyli dobór metody tłumaczenia dostosowany do funkcji tekstu oraz sytuacji jego użycia.

W literaturze przedmiotu, zwłaszcza tej powstałej w kręgu niemieckojęzycznym, często podkreślano trudności związane z tłumaczeniem frazeologizmów oraz brak możliwości osiągnięcia pełnej ekwiwalencji stylistycznej, semantycznej i kulturowej: *Man spricht von der Unmöglichkeit des Übersetzens und vom Scheitern, vom Ungenügen jeder Übersetzung* (Koller 1972: 9).

Choć rzeczywista nieprzekładalność należy do rzadkości, różnice między oryginałem a przekładem wynikają zazwyczaj z odmienności strukturalnych, stylistycznych i kulturowych między językami. W nowszych ujęciach rezygnuje się z postrzegania nieprzekładalności jako cechy właściwej frazeologizmom, skupiając się raczej na typologii ekwiwalentów i funkcji tłumaczonego wyrażenia.

Jak zauważa Kaldiewa-Zahariewa, twierdzenie to należy uznać za nieadekwatne wobec złożonych mechanizmów przekładu (Калдиева-Захариева 2013: 64).

Podzielając ten pogląd, warto przywołać rozróżnienie na dwa pozostałe typy ekwiwalencji: ekwiwalencję systemową i ekwiwalencję przekładową (Blagoeva, Sosnowski 2022: 51). Pierwsza z nich dotyczy relacji pomiędzy jednostkami należącymi do różnych systemów językowych i stanowi główny przedmiot badań frazeografii teoretycznej oraz praktycznej. Druga natomiast odnosi się do przekładu tekstowego i zakłada konieczność osiągnięcia funkcjonalnej odpowiedniości między jednostką oryginału a jej odpowiednikiem w języku docelowym. To rozróżnienie okazuje się szczególnie istotne w kontekście tłumaczenia frazów nacechowanych pragmatycznie i sytuacyjnie, których funkcja w tekście często przeważa nad ich wartością systemową.

Powracając do problematyki przekładu reproduktów wielowyrazowych, warto podkreślić, że szczególnie w przypadku jednostek o charakterze emotywnym i fatycznym rzadko możliwe jest wskazanie jednego, w pełni równoważnego odpowiednika. Częściej tłumacz wybiera spośród kilku wyrażen w języku docelowym, które przekazują ten sam sens, ale różnią się strukturą formalną, stopniem idiomatyczności lub zakorzenieniem w danym rejestrze językowym. Taka wielowariantowość przekładu wymaga nie tylko ich identyfikacji, lecz także gradacji ekwiwalentów, czyli określenia stopnia przystawalności danej jednostki do oryginału (Gasek 2017: 147). Dobór ekwiwalentu powinien uwzględniać nie tylko zgodność semantyczną, ale także częstość użycia i typowe konteksty. Ustalenie ich reprezentatywnej i

inwariantywnej formy, szczególnie w odniesieniu do jednostek nowych lub potocznych, często wymaga odwołania do aktualnych danych korpusowych.

W tym kontekście szczególne znaczenie ma pytanie Andrzeja Bogusławskiego o to, „jakiego zdania w kodzie K” użyłby w sposób maksymalnie naturalny ‘rodzimy’ (czy też ‘rodowity’) nosiciel K” (tzw. native speaker) w sytuacji właściwej zdaniu wyjściowemu w kodzie K”?” (Bogusławski 1997: 14).

Jak zauważa Z. Wawrzyniak, zadaniem tłumacza jest stworzenie takiego odpowiednika w języku docelowym, który odpowiada tekstowi wyjściowemu pod względem komunikacyjno-semantycznym (Wawrzyniak 2002: 41). Autor rozwija tę myśl, pisząc:

„Wytworzenie translatu jest każdorazowo utrwaleniem rezultatów rozumienia tekstu wyjściowego. Oryginał i translat, jako dwa teksty różnojęzyczne, mają swoje *tertium comparationis* w tłumaczu, a dokładniej mówiąc, w jego rezultatach rozumienia, osiągniętych w odniesieniu do oryginału i utrwalonych w sformułowaniach składających się na translat” (Józwiak-Dądela 2007: 46).

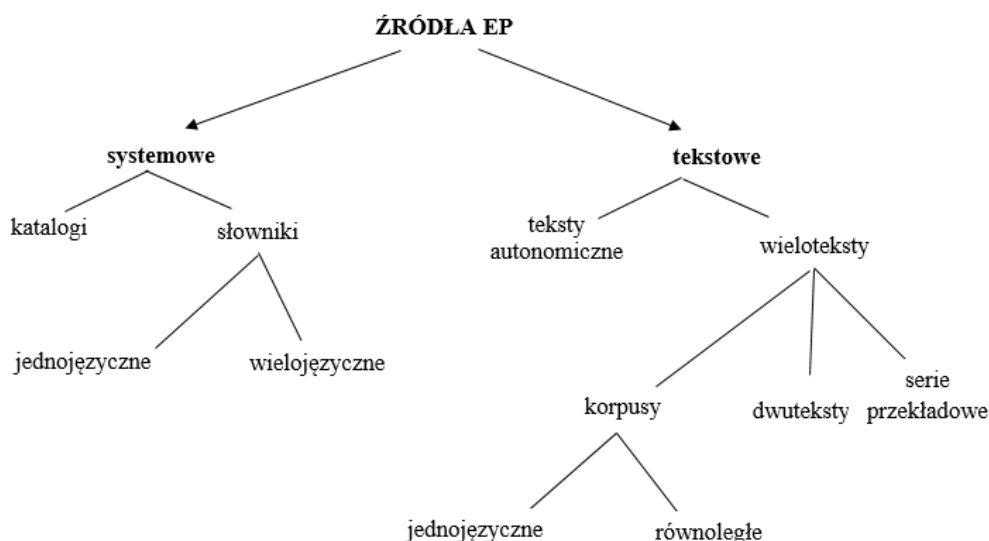
Bardzo przydatne okazują się pojęcia *translandu* i *translatu*, odnoszące się odpowiednio do fragmentu tekstu oryginalnego oraz odpowiadającego mu odcinka tekstu przekładu. Zależność ta nie ogranicza się do wyrazów – może odnosić się także do większych jednostek, takich jak frazy czy zdania. Ich relacja podkreśla, że przekład dotyczy nie abstrakcyjnych elementów systemu językowego, lecz konkretnych tekstów funkcjonujących w określonych aktach komunikacyjnych. W takim rozumieniu tłumaczenie odbywa się zatem na poziomie tekstu jako jednostki aktualnej, osadzonej w kontekście i służącej rzeczywistej komunikacji. Kluczowe staje się zatem pytanie o to, gdzie i w jaki sposób należy poszukiwać adekwatnych ekwiwalentów tłumaczeniowych dla jednostek o charakterze emotywnym i fatycznym.

Odpowiedzi na postawione pytanie dostarcza koncepcja Wojciecha Chlebdy, według którego poszukiwanie odpowiedników przekładowych jest *tropieniem*, tzn. procesem sytuacyjnym, otwartym i z założenia nieomkniętym. W słowniku przekładowym nie sposób wskazać raz na zawsze ustalonego zestawu translatów, ponieważ relacja ekwiwalencji zależy od kontekstu, aktualności użycia i funkcji komunikacyjnej. Najważniejszym staje się zatem uchwycenie kryterium sytuacyjności i wyznaczenie określonego *tertium comparationis*. W ujęciu W. Chlebdy ekwiwalenty przykładowe, czyli praktyczne eksponenty relacji ekwiwalencji, mogą być tropione według następujących zasad:

1. Głównym źródłem odpowiedników są teksty – wypowiedzi w języku docelowym, formułowane przez jego rodzimych użytkowników w konkretnych sytuacjach życiowych;

2. Za materiał porównawczy mogą służyć zarówno teksty autonomiczne, jak i tzw. *dwuteksty* i *wieloteksty*, czyli różnojęzyczne realizacje tego samego tekstu wyjściowego;
3. Kierunek poszukiwań może być klasyczny – od jednostki w tekście źródłowym do tekstu docelowego – lub odwrotny: od tekstu w języku obcym do rodzimego odpowiednika;
4. Poszukiwanie ekwiwalentu przykładowego powinno uwzględniać określone *tertium comparationis* (czyli podstawę porównania);
5. Jeśli dana jednostka docelowa już funkcjonuje jako ekwiwalent, warto sprawdzić, czy odpowiadająca jej jednostka wyjściowa rzeczywiście pokrywa się funkcjonalnie i znaczeniowo. (Chlebda 2011a: 15).

Z kolei pojęcie ekwiwalencji w ujęciu Chlebdy może mieć charakter tekstowy (między tekstami) lub słownikowy (między jednostkami języka), przy czym oba podejścia są komplementarne: dane słownikowe opierają się na realizacjach tekstowych, a analiza tekstów pozwala wyodrębniać nowe ekwiwalenty. W tym kontekście źródła ekwiwalentów przykładowych można podzielić na *systemowe* (językowe) i *tekstowe* (oparte na korpusach), co znajduje odzwierciedlenie w poniższym schemacie (Chlebda 2011b: 26):

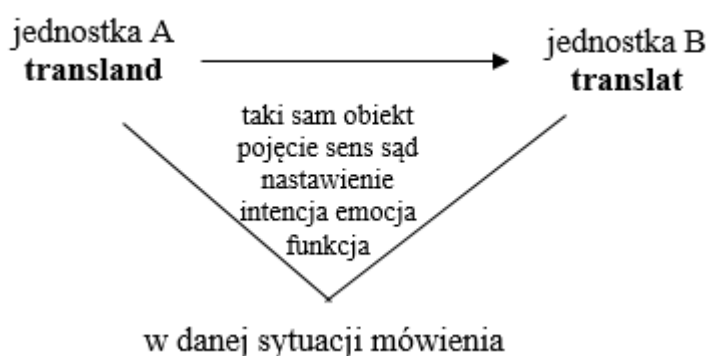


Przedstawiony schemat ilustruje dwupodział źródeł ekwiwalentów przykładowych (EP), w którym lewa strona (źródła systemowe) zawiera zasoby językowe niezależne od kontekstu komunikacyjnego – takie jak katalogi czy słowniki – w których jednostki języka funkcjonują jako abstrakcyjne reprezentacje znaczeń. Drugą kategorię stanowią źródła tekstowe, czyli zbiory rzeczywistych wypowiedzi i struktur tekstowych, takie jak teksty

autonomiczne oraz bardziej złożone układy, tj. wieloteksty. Do ich najczęstszych form należą korpusy (jedno- i wielojęzyczne), dwuteksty (oryginał i przekład) oraz serie przekładowe.

W kontekście wspomnianego *tertium comparationis* należy podkreślić, iż powinno ono odnosić się do obiektu, pojęcia lub sytuacji, przy czym – jak zauważa Ewa Białek – istotą poszukiwania ekwiwalentu tłumaczeniowego jest dążenie do odnalezienia jednostki o zbliżonym sensie (Białek 2009: 41).

Z kolei Tadeusz Piotrowski podkreśla, że rozumienie *tertium* jako „sytuacji” należy traktować raczej jako wynik intuicyjnej konceptualizacji, dokonywanej przez tłumacza, który identyfikuje i wyodrębnia cechy relewantne z punktu widzenia przekładu (Piotrowski 2011b: 184). Zestawienie translatu i translandu opiera się na założeniu, że jednostki te można uznać za ekwiwalentne, o ile odnoszą się do tych samych treści – takich jak obiekt, pojęcie, sąd, intencja czy funkcja – w ramach konkretnej sytuacji komunikacyjnej. Według tego ujęcia należy również poszukiwać odpowiedników dla jednostek o charakterze emotywnym i fatycznym, których znaczenie i funkcja ujawniają się przede wszystkim w kontekście użycia. Poniższy schemat obrazuje kluczowe kryteria w kontekście danej sytuacji mówienia (Chlebda 2011b: 28):



W oparciu o analizę różnych źródeł ekwiwalentów przekładowych przy zastosowaniu kryterium sytuacyjności można przyjąć istnienie kontinuum relacji między jednostką wyjściową a jej odpowiednikiem w języku docelowym. Relacje te obejmują cztery podstawowe typy ekwiwalencji słownikowej: pełną, zachwianą, zerową oraz funkcjonalną. Ich granice są płynne, a poszczególne typy przechodzą jeden w drugi, tworząc spektrum rozwiązań o różnym stopniu odpowiedniości.

Ekwiwalenty pełne cechuje zgodność znaczeniowa, obrazowa, pragmatyczna, stylistyczna, a nierzadko również gramatyczna.

Ekwiwalenty zachwiane odwzorowują jednostkę wyjściową tylko częściowo, co skutkuje asymetrią treściową i pragmatyczną.

W przypadku *ekwiwalentów zerowych* w języku docelowym brak ustalonego odpowiednika, co wiąże się z koniecznością zastosowania strategii kompensacyjnych, takich jak opisy, parafrazy, kalki czy metakomunikaty.

Na tle tych rozwiązań szczególne znaczenie – zwłaszcza w odniesieniu do jednostek o charakterze emotywnym i fatycznym – zyskuje ekwiwalencja funkcjonalna, w której podstawowym kryterium przekładowym staje się wspólna funkcja komunikacyjna i zbliżony sens, niezależnie od różnic w strukturze formalnej, stylu czy obrazowości. W świetle powyższych rozważań w dalszej części rozdziału podjęta zostanie próba identyfikacji ekwiwalentów tłumaczeniowych dla jednostek fatycznych i emotywnych w układzie polsko-rosyjskim i polsko-czeskim.

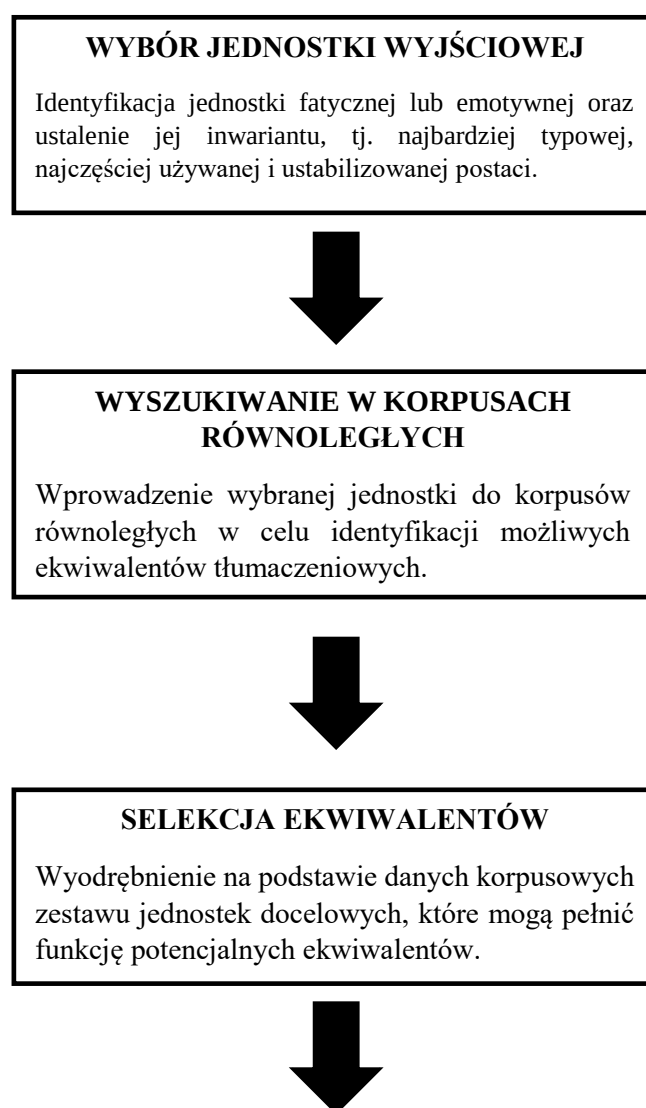
Analiza będzie oparta na wybranych źródłach systemowych, z uwzględnieniem kryterium sytuacyjności oraz dążenia do uzyskania identycznej funkcji komunikacyjnej odnalezionych jednostek. W tym ujęciu *tertium comparationis* – czyli podstawą porównania jednostek oryginału i przekładu – będzie zgodność sytuacyjna oraz równoważność funkcji pragmatycznej, rozumianej jako zdolność jednostki docelowej do wywołania u odbiorcy zbliżonego efektu komunikacyjnego. Takie podejście, zgodne z koncepcją tropienia ekwiwalentów, zakłada wykorzystanie komplementarnych metod analitycznych i pozwala na elastyczne dopasowanie ekwiwalentów do ich rzeczywistego użycia w różnych sytuacjach komunikacyjnych.

7.1 Metoda korpusowa

Pierwsza z zaprezentowanych metod ustalania ekwiwalentów tłumaczeniowych dla wybranych jednostek emotywnych i fatycznych języka polskiego polega na wykorzystaniu korpusów równoległych. W ujęciu tym będą one pełnić funkcję narzędzia wyjściowego, stanowiącego punkt odniesienia dla dalszych etapów tropienia odpowiedników przekładowych. Z uwagi na specyfikę badanych jednostek – silnie osadzonych w kontekście sytuacyjnym i nacechowanych pragmatycznie – należy przyjąć zasadę ograniczonego zaufania wobec danych uzyskiwanych z korpusów. Innymi słowy, propozycje przekładowe wygenerowane przez narzędzia korpusowe nie powinny być traktowane jako ostateczne, lecz wymagają dodatkowej weryfikacji w innych źródłach. Jak już wspomniano, przekład reproduktów rzadko umożliwia ustalenie ścisłej relacji jeden do jednego, a poszukiwanie adekwatnych ekwiwalentów stanowi proces otwarty, rozproszony i trudny do ujęcia w ramach jednolitej procedury. Potencjalne odpowiedniki mogą być rozproszone w różnych źródłach systemowych – zarówno

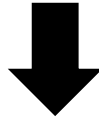
papierowych, jak i cyfrowych, zawierających źródła internetowe, słowniki, korpusy jednojęzyczne, materiały internetowe oraz audiowizualne.

Kryterium nadrzędnym staje się zatem zgodność sytuacyjna oraz spójność pragmatyczna jednostek rozumiana jako zachowanie sensu i funkcji komunikacyjnej wyrażenia źródłowego w analogicznych kontekstach w języku docelowym. Na potrzeby części analitycznej i w celach egzemplifikacyjnych, wybrano przykłady emotywów i fatyzmów języka polskiego, dla których podjęta zostanie próba odnalezienia możliwie najbardziej zbliżonych ekwiwalentów w języku rosyjskim i czeskim. W celu przejrzystego zobrazowania kolejnych etapów proponowanej procedury, cały proces został ujęty w formie następującego schematu:



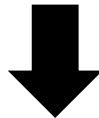
**ANALIZA SEMANTYCZNA I PRAGMATYCZNA
W KORPUSACH JEDNOJĘZycznych**

Wprowadzenie wyodrębnionych ekwiwalentów do odpowiednich korpusów jednojęzycznych. Analiza uzyskanych konkordancji w celu określenia znaczenia oraz funkcji pragmatycznej.



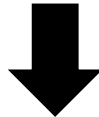
OKREŚLENIE FREKWENCJI

Określenie frekwencji jednostek w oparciu o dane ilościowe uzyskane z korpusów jednojęzycznych.



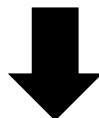
**WERYFIKACJA W DODATKOWYCH
ŹRÓDLACH**

Włączenie danych pozakorpusowych pochodzących z różnorodnych źródeł leksykograficznych i internetowych



**ANALIZA PORÓWNAWCZA I USTALENIE
PAR PRZEKŁADOWYCH**

Na podstawie danych z korpusów i źródeł pomocniczych ustala się potencjalne pary przekładowe: polsko-rosyjskie i polsko-czeskie



**WERYFIKACJA PRZEZ PRZEKŁAD
ODWROTNY**

Zastosowanie przekładu odwrotnego (ros.–pl., cz.–pl.,) w celu oceny relacji ekwiwalencyjnej i weryfikacji wcześniejszych założeń.

Innymi słowy, procedura rozpoczyna się od wyboru jednostki wyjściowej w języku polskim – wyrażenia fatycznego lub emotywnego. Na tym etapie dokonuje się określenia jego częstości występowania na podstawie danych pozyskanych z zasobów korpusowych takich jak Narodowy Korpus Języka Polskiego (NKJP), Araneum Polonicum Maius (będący częścią zbioru NKJCz) oraz polska część korpusu InterCorp v16ud (w momencie analizy korpus MoncoPL był niedostępny). Równolegle ustalany jest *invariant* danej jednostki, tj. jej najczęściej występująca i najbardziej ustabilizowana forma. Dalsza analiza uwzględnia również rozpoznanie wszystkich możliwych znaczeń jednostki w kontekście użycia. Następnie jednostka ta zostaje wprowadzona do odpowiednich korpusów równoległych – w niniejszym badaniu są to InterCorp v16ud oraz równoległy korpus NKJR – w celu identyfikacji potencjalnych ekwiwalentów tłumaczeniowych w językach docelowych. Spośród uzyskanych wyników wyodrębnia się najbardziej odpowiednie propozycje przekładowe, które zostają poddane dalszej weryfikacji w korpusach jednojęzycznych. W przypadku języka rosyjskiego wykorzystane zostaną Araneum Russicum Maius, rosyjska część InterCorp v16ud oraz zasoby korpusu NKJR, natomiast dla języka czeskiego – czeska część InterCorp v16ud, Araneum Bohemicum Maximum oraz korpus ORAL v1. Analiza konkordancji umożliwi określenie znaczenia i funkcji pragmatycznych poszczególnych jednostek oraz oszacowanie ich frekwencji.

Kolejny etap obejmuje weryfikację pozyskanych danych w źródłach pozakorpusowych, takich jak słowniki jednojęzyczne i przekładowe, translatory internetowe, fora dyskusyjne oraz inne zasoby internetowe. W tym momencie rozpoczyna się otwarty i rozproszony proces tropienia ekwiwalentów, którego przebieg może być trudny do jednoznacznego przewidzenia, ze względu na rozproszenie danych językowych. Po zebraniu, potwierdzeniu i uporządkowaniu potencjalnych par przekładowych (polsko-rosyjskich i polsko-czeskich), przeprowadza się etap przekładu odwrotnego (ros.–pl., cz.–pl.). Jego celem jest ostateczna weryfikacja relacji ekwiwalencyjnych oraz potwierdzenie słuszności wcześniej sformułowanych założeń.

Emotywy

W pierwszej kolejności przeprowadzona zostanie analiza mająca na celu ustalenie odpowiedników tłumaczeniowych w języku rosyjskim i czeskim dla wybranego emotywu języka polskiego.

Pierwsza z nich kwalifikowana jest jako emotywy reakcyjny, ogólnie wyraża postawę obojętności i przyjmuje postać wyrażenia *rób, co chcesz*. Przeprowadzone uprzednio

obserwacje i analizy pozwalają stwierdzić, że spośród pozostałych wariantów takich jak *rób sobie, co chcesz; rób, jak chcesz* czy *rób se, co chcesz*, za najbardziej rozpowszechnioną należy uznać wskazaną przez nas formę, która w dalszej analizie traktowana będzie jako inwariantywna. Weryfikacja częstości występowania w NKJP, Araneum Polonicum Maius (APM) oraz polskiej części korpusu InterCorp v16ud wykazała łącznie 1328 poświadczeń. Do zestawienia włączono również te konkordancje, w których pominięto przecinek, o ile zachowany został sens jednostki (NKJP – 134 poświadczenia, APM – 215, InterCorp – 979). Przejdźmy zatem do ustalenia możliwych znaczeń analizowanej jednostki. Na podstawie uzyskanych konkordancji można stwierdzić, że *rób, co chcesz* realizuje się przede wszystkim w dwóch podstawowych funkcjach pragmatycznych:

EMOTYW	ZNACZENIE	PRZYKŁAD UŻYCIA
1. <i>rób, co chcesz</i>	Wyraz rzeczywistej obojętności. Sygnalizuje brak zaangażowania emocjonalnego nadawcy oraz przekazanie odbiorcy pełnej swobody działania. Nadawca rezygnuje z wpływu na sytuację i nie zamierza ingerować w wybory rozmówcy.	<i>Andrzej wyjął ze starej skórzanej torby notatnik i dyktafon. - Pozwolisz, że będę notował? - niespodziewanie przeszedł z Niemcem na ty. - <u>Rób, co chcesz.</u> Opowiem ci, jak to wyglądało, ale w twoim kraju pewnie nikt ci nie uwierzy.</i>
2. <i>rób, co chcesz</i>	Pozorna zgoda ukrywająca niezadowolenie, rozczarowanie lub presję emocjonalną. Wyrażenie jedynie formalnie dopuszcza swobodę decyzji, lecz w rzeczywistości towarzyszy mu rozczarowanie, ironia lub ukryta dezaprobata. Wypowiedź może pełnić funkcję manipulacyjną, sugerując, że nadawca oczekuje innego rozwiązania.	<i>– Wcale się nie całowaliśmy.[...] Fala mnie na nią rzuciła, co miałem robić? – Fala cię rzuciła? Akurat ciebie i akurat w jej ramiona? Już mi takich bzdur nie opowiadaj. A zresztą nie musisz się tłumaczyć, <u>rób, co chcesz.</u></i>

Jak można zauważyć, analizowany reprodukt jako replika może wyrażać zarówno rzeczywistą obojętność nadawcy, jak i ukryte niezadowolenie z decyzji rozmówcy, często implikujące oczekiwanie, że ten postąpi inaczej. Konkordancje wskazują na częstsze występowanie drugiego z tych znaczeń, co sugeruje jego dominującą realizację w rzeczywistej komunikacji. Z tego względu w procesie przekładu konieczne jest uwzględnienie wieloznaczności danej jednostki, ponieważ jej odpowiednik w języku rosyjskim lub czeskim

nie musi odzwierciedlać wszystkich funkcji znaczeniowych translandu. Kolejnym etapem procedury jest wprowadzenie wyrażenia *rób, co chcesz* do:

1. wyszukiwarki korpusu równoległego polsko-rosyjskiego InterCorp, z zastosowaniem atrybutu *word*, który pozwala zachować dokładną pisownię jednostki wyjściowej. W tym przypadku uwzględniane są również poświadczenia z pominięciem przecinka w strukturze emotywu;

2. polsko-rosyjskiej części korpusu równoległego NKJR, poprzez okno wyszukiwawcze „Поиск точных форм”. Poniżej zaprezentowano przykłady konkordancji zawierające najbardziej zbliżone odpowiedniki tłumaczeniowe, a także zestawienie potencjalnych ekwiwalentów wyodrębnionych z obu korpusów, wraz z informacją o ich częstotliwości występowania konkretnych realizacji par przekładowych:

INTERCORP

☐ Puskin-Kapitanska_dcer	Idź sobie na wszystkie cztery wiatry i rób, co chcesz .	Puskin-Kapitanska_dcer	Ступай себе на все четыре стороны и делай что хочешь.
☐ Ajtmatov-Popraviste	Idź i rób, co chcesz .	Ajtmatov-Popraviste	Иди, делай свое дело!
☐ Waltari-EgyptanSinuhet	Ale na Amona, graj w kości, trwój złoto na kobiety, pij do upadłego, rób co chcesz , bylebyś tylko nie skazywał się na zagładę szalonymi słowami, bo kocham cię wielce, panie mój, Sinuhe, i nie chcę, żeby ci się przytrafiło coś złego.	Waltari-EgyptanSinuhet	Ради Амона, играй в кости, трать золото на женщин, упивайся допьяна! Делай что угодно, только не проводи время в пустых рассуждениях! Говорю так, потому что всем сердцем люблю тебя, мой господин, и не хочу, чтобы с тобой стряслась беда.
☐ _SUBTITLES	Rób, co chcesz .	_SUBTITLES	Делай как хочешь.

Ekwiwalent	Liczba poświadczeń danego ekwiwalentu	Procentowy udział w ogólnej liczbie poświadczeń
<i>делай, что хочешь</i>	140 poświadczeń	32,7%
<i>делай, как хочешь</i>	12 poświadczeń	2,8%
<i>делай, что угодно</i>	7 poświadczeń	1,6%
<i>делай своё дело</i>	2 poświadczenia	0,4%

NKJR

1. Stanisław Lem. Solaris (1961) | Станислав Лем. Солярис (Дм. Брускин, 1973) ①

польский:

Ruściłem go. - *Rób, co chcesz*. - Chodzi o to, że Sartorius będzie usiłował ukryć pewne fakty. 🗣️ ↔️

русский:

Я отпустил его. — Делай что хочешь. — Дело в том, что Сарториус постарается скрыть некоторые факты. 🗣️ ↔️

Ekwiwalent	Liczba poświadczeń danego ekwiwalentu	Procentowy udział w ogólnej liczbie poświadczeń
<i>делай, что хочешь</i>	6 poświadczeń	85,7%

Wyszukanie odpowiedników tłumaczeniowych w obu korpusach równoległych wykazało, że potencjalnie najczęstszym rosyjskim odpowiednikiem jest dosłowne tłumaczenie *делай, что хочешь*. Ogólna liczba poświadczeń była stosunkowo niska — łącznie uzyskano jedynie 146 par przekładowych z obu źródeł. Zarejestrowano również nieliczne bądź pojedyncze wystąpienia innych jednostek, takich jak *Делай, как хочешь*; *Делай своё дело* oraz *Делай, что угодно*. Ich incydentalna frekwencja nie powinna jednak stanowić podstawy do ich wykluczenia, gdyż każda z nich wymaga dalszej weryfikacji zarówno pod względem znaczeniowym, jak i frekwencyjnym. W związku z tym wszystkie wyodrębnione jednostki zostaną poddane analizie we wspomnianych korpusach jednojęzycznych, tj. Araneum Russicum Maius, rosyjskiej części InterCorp v16ud oraz w zasobach korpusu NKJR. Poniżej przedstawiono znaczenia wyodrębnionych jednostek, ustalone na podstawie analizy konkordancji, zilustrowane przykładami użycia, a także łączną liczbę ich wystąpień w wybranych korpusach jednojęzycznych:

EKWIWALENT	ZNACZENIE	PRZYKŁAD UŻYCIA
<i>Делай, что хочешь</i>	1. Wyraz rzeczywistej obojętności. Wyrażenie sygnalizuje brak emocjonalnego zaangażowania nadawcy i pozostawienie rozmówcy pełnej swobody działania. 2. Pozorna zgoda ukrywająca niezadowolenie lub presję emocjonalną. Fraza formalnie dopuszcza swobodę decyzji, lecz w rzeczywistości towarzyszy jej rozczarowanie, ironia lub ukryta dezaprobata. Może służyć jako forma biernego oporu lub pasywnej manipulacji.	1. Так вот, что мы будем делать с тобой? Ты для меня помеха. Как и эта антикварная рухлядь Доктора. Так, дай мне подумать. - Мне уже все равно. <u>Делай, что хочешь.</u> Просто давай покончим с этим. 2. Ты не наш командир – Но ты будешь делать то, что я скажу! Жан Ропер огляделся, напрасно ища поддержки среди сидящих в подвале. – <u>Делай, что хочешь,</u> – сказал он уныло.

<i>Делай, как хочешь</i>	Neutralna zgoda na sposób działania. Wyrażenie sygnalizuje akceptację dla wyboru sposobu wykonania danej czynności, pozostawiając odbiorcy swobodę decyzji, jak coś zrobić. Nie zawiera ukrytej dezaprobaty czy ironii.	- Так ты хочешь, чтобы я ушёл или остался? <i>Определись уже, куколка.</i> - Мне всё равно, <u>делай, как хочешь.</u>
<i>Делай своё дело</i>	Zachęta do wykonywania własnych obowiązków. Wyrażenie sugeruje koncentrację na zadaniach przypisanych mówiącemu i unikanie ingerencji w działania innych.	Мы этим занимаемся, брат. Завтра, у Джо встреча с Максимо, он попробует сбить цену. - Хорошо. <u>Делай своё дело</u> и держи нас в курсе.
<i>Делай, что угодно</i>	Pozorna zgoda ograniczana warunkiem lub zastrzeżeniem. Wyrażenie formalnie dopuszcza pełną swobodę działania, jednak często towarzyszy mu dalsze ograniczenie lub ukryta dezaprobata.	Следуй за ним, куда бы ни пошёл. И <u>делай, что угодно</u> , лишь бы мой мужчина не погиб снаружи.

EKWIWALENT	FREKWENCJA
<i>Делай, что хочешь</i>	1272 poświadczenia
<i>Делай своё дело</i>	129 poświadczeń
<i>Делай, как хочешь</i>	128 poświadczeń
<i>Делай, что угодно</i>	51 poświadczeń

Przeprowadzone analizy pozwalają stwierdzić, że spośród potencjalnych odpowiedników tłumaczeniowych polskiego wyrażenia *rób, co chcesz* największą zgodność znaczeniową i funkcjonalną wykazuje *делай, что хочешь* (1272 poświadczenia). Jednostka ta realizuje dwie główne funkcje pragmatyczne: wyrażenie rzeczywistej obojętności oraz pozornej zgody z ukrytą dezaprobatą. Należy zaznaczyć, że również ta forma często pojawia się w strukturach modyfikowanych przez wyrażenia ograniczające (np. *но, только, чтобы*). Na uwagę zasługuje również znaleziona konkordancja: *-Что мне делать? — спросила девица, сидя на диване. Делай что хочешь!...* — хором заявили мы с Александром, której znaczenie może być bliższe wyrażeniu *rób, co хочешь* bądź *róб, co зechcesz*. Wśród pozostałych translatów bliskie podobieństwo wykazuje *делай, как хочешь* (128 poświadczeń), który odnosi się przede wszystkim do sposobu działania i wyraża zgodę na formę wykonania czynności. Z tego względu jednostka ta może być uznana za ekwiwalent częściowy, szczególnie w

sytuacjach nacechowanych neutralnie lub wtedy, gdy podkreślany jest sposób, a nie treść działania. Zdecydowanie niższą frekwencję odnotowano w przypadku wyrażenia *делай, что угодно* (51 poświadczeń), które – mimo formalnej bliskości do *делай, что хочешь* – występuje głównie w strukturach warunkowych. Typowe są dla niej konstrukcje ograniczające swobodę działania, np. *делай, что угодно, только не...; лишь бы...; главное, чтобы*. Tym samym fraza ta realizuje funkcję reakcyjną zbliżoną do *делай, что хочешь*, lecz nie oddaje pełnego zakresu znaczeniowego wyrażenia wyjściowego. Z kolei jednostka *делай своё дело* (129 poświadczeń) sygnalizuje potrzebę skupienia się na własnych obowiązkach i rezygnacji z ingerencji w działania innych. W odróżnieniu od wcześniej omówionych jednostek, nie realizuje żadnej z funkcji wyrażenia *рób, co хочеш*, przez co nie może być traktowana jako jego odpowiednik przekładowy. Podsumowując, analizowany materiał wskazuje jednoznacznie, że *делай, что хочешь* stanowi najbardziej prawdopodobny odpowiednik polskiego *рób, co хочеш*, zarówno pod względem leksykalnym, jak i funkcjonalnym.

Końcowym etapem poprzedzającym ustanowienie adekwatnej pary przekładowej pozostaje proces tropienia, polegający na weryfikacji każdej z wyodrębnionych jednostek w źródłach pozakorpusowych. Ze względu na ich otwarty i nieprzewidywalny charakter nie sposób w tym momencie określić, czy dana jednostka zostanie odnotowana w określonym zasobie, ani jakie informacje mogą się tam znaleźć. Poszukiwania te mają zatem charakter rozproszony i uwzględniają przede wszystkim słowniki jednojęzyczne, translatory automatyczne, a także różnorodne źródła internetowe, takie jak artykuły językoznawcze, fora dyskusyjne oraz strony poświęcone językowi rosyjskiemu. W dalszej części zaprezentowano zestawienie jednostek, dla których odnaleziono konkretne informacje w wybranych źródłach pozakorpusowych, zgodnie z układem przedstawionym poniżej:

<i>Делай, что хочешь</i>		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Google translate	Fraza przetłumaczona jako „ <i>рób, co хочеш</i> ”.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Pragmaticon.ru https://pragmaticon.ruscorpora.ru/exp/94 [data dostępu: 05.05.2025].	Rejestracja jednostki jako wyrażenia, realizującego przede wszystkim funkcję obojętności oraz negatywnej oceny działań rozmówcy.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Artykuł psychologiczny. https://www.psychologies.ru/artic	Artykuł interpretuje frazę <i>делай, что хочеш</i> jako wypowiedź, która – mimo	Dominującą funkcją jednostki jest wyrażenie zdenerwowania

les/chto-na-samom-dele-znachit-fraza-delai-chto-khochesh-skrytye-smysly-i-poslaniya/ [data dostępu: 05.05.2025].	pozornej neutralności – często pełni funkcję pasywno-agresywną, wyrażając emocjonalny dystans, rozczarowanie lub ukrytą dezaprobatę.	oraz ukrytego oddziaływania na odbiorcę. Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Artykuł https://dzen.ru/a/ZGB54IEDcnnUTguL [data dostępu: 05.05.2025].	Artykuł interpretuje frazę <i>делай, что хочешь</i> jako wypowiedź manipulacyjną, stosowaną podczas kłótni.	Dominującą funkcją jednostki jest wyrażenie zdenerwowania oraz ukrytego oddziaływania na odbiorcę. Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>Делай, как хочешь</i>		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Glosbe	Fraza przetłumaczona jako „ <i>rób, jak chcesz</i> ”.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Sinonim.org sinonim.org/s/делай как хочешь [data dostępu: 05.05.2025].	Wskazanie synonimów dla danej jednostki, takich jak: <i>решиай сам; делай по-своему; как считаешь нужным</i> .	Wyrażenie neutralnie nacechowane. Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Forum internetowe https://otvet.mail.ru/question/64636040?utm [data dostępu: 05.05.2025].	Fraza interpretowana przez użytkowników jako ukryta manipulacja.	W odróżnieniu od wcześniejszych obserwacji fraza została uznana za wyrażenie wyłącznie nacechowane emocjonalnie i manipulacyjne.

<i>Делай, что угодно</i>		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Glosbe	Fraza przetłumaczona jako „ <i>rób, co chcesz</i> ”. Dane uzyskane z analizy konkordancji potwierdzają współwystępowanie ograniczników.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Artykuł eseistyczny https://syg.ma/@affabricatus/dielai-chto-ughodno-mnoghotochiie [data dostępu: 5.05.2025].	Fraza stosowana jako wyraz absolutnej wolności działania („ <i>rób, co chcesz</i> ”) w strukturach typu <i>делай что угодно, лишь бы... / чтобы...</i> , co wskazuje na obecność ograniczników modyfikujących pierwotny sens.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>Делай своё дело</i>		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Google Translate	Fraza przetłumaczona jako „wykonaj swoją pracę”.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Kartaslov.ru kartaslov.ru/значение-слова/делать+своё+дело [data dostępu: 05.05.2025].	Wyrażenie występuje jako element idiomu „Мавр сделал своё дело, мавр может уходить”, który oznacza osobę, której zadanie zostało wykonane i której obecność przestaje być potrzebna	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Forum internetowe. https://otvet.mail.ru/question/166912763?utm [data dostępu: 05.05.2025].	Wyrażenie rozumiane jako wykonywanie własnych obowiązków w sposób konsekwentny, spokojny i niezależny od okoliczności, bez ingerencji w sprawy innych.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

Przeprowadzona analiza danych pozakorpusowych potwierdza wcześniejsze obserwacje oparte na materiałach korpusowych. Wszystkie zgromadzone źródła wskazują, że najbardziej trafnym odpowiednikiem tłumaczeniowym polskiego reproduktu *rób, co chcesz* jest rosyjska jednostka *делай, что хочешь*. Wyrażenie to można uznać za ekwiwalent pełny, ponieważ – poza zgodnością leksykalną – pełni ono te same funkcje pragmatyczne, tj. zarówno wyrażenie rzeczywistej obojętności nadawcy, jak i pozornej zgody, nacechowanej negatywnie i zawierającej element ukrytego oddziaływania na odbiorcę. Na tej podstawie uzasadnione jest przyjęcie, że *делай, что хочешь* stanowi najpełniejszy odpowiednik analizowanej jednostki wyjściowej. W dalszej kolejności, jako jednostkę o nieco niższej frekwencji, ale zbliżoną znaczeniowo i strukturalnie, należy wskazać *делай, как хочешь*. Choć wyrażenie to interpretowane jest przeważnie jako neutralna zgoda na sposób działania, dane pozakorpusowe sugerują możliwość jego użycia także w funkcji nacechowanej emocjonalnie, zbliżonej do polskiego wyrażenia *rób, jak chcesz*. Pewne podobieństwo funkcjonalne wykazuje również fraza *делай, что угодно*, której znaczenie może pokrywać się z analizowaną jednostką wyjściową głównie w obecności struktur ograniczających swobodę działania (np. *только не...; лишь бы...; чтобы...*). Niska frekwencja oraz warunkowe użycie sprawiają jednak, że wyrażenie to należy uznać za ekwiwalent najmniej odpowiedni do wyrażenia *róб, co chcesz*.

Natomiast jednostka *делай своё дело* odnosi się do zupełnie innego typu sytuacji komunikacyjnej. Służy ona wskazaniu konieczności wykonywania własnych obowiązków i nie realizuje żadnej z funkcji znaczeniowych jednostki wyjściowej, przez co nie może być traktowana jako jej odpowiednik tłumaczeniowy. Uwzględniając powyższe obserwacje, możliwe jest zaproponowanie gradacji potencjalnych par przekładowych, uwzględniającej stopień ich adekwatności wobec translandu.

Najwyższy poziom ekwiwalencji wykazuje para: 1. *rób, co chcesz* → *делай, что хочешь*, następnie 2. *rób, co chcesz* → *делай, как хочешь*, natomiast na trzecim miejscu należy umieścić parę 3. *rób, co chcesz* → *делай, что угодно*.

Końcowy etap analizy przewiduje zastosowanie przekładu odwrotnego, przeprowadzonego na podstawie danych z rosyjsko-polskiej części korpusu InterCorp v16ud oraz równoległego korpusu NKJR. W ramach tego etapu wcześniej wyodrębnione jednostki rosyjskie traktowane są jako translandy, wobec których identyfikowane są ich polskie odpowiedniki. W dalszej części przedstawiono zestawienie tych odpowiedników wraz z danymi dotyczącymi ich frekwencji w materiale korpusowym:

TRANSLAND ROSYJSKI	TRANSLAT POLSKI	FREKWENCJA
<i>делай, что хочешь</i>	<i>rób, co chcesz</i>	115 poświadczeń (37%)
	<i>możesz robić, co chcesz</i>	8 poświadczeń (2,6%)
<i>делай, как хочешь</i>	<i>rób, co chcesz</i>	5 poświadczeń (16%)
<i>делай, что угодно</i>	<i>zrób cokolwiek</i>	4 poświadczenia (12,5%)

Analiza przekładu odwrotnego wykazała, że najczęściej wskazywanymi i zarazem najbardziej trafnymi odpowiednikami dla polskiego re produktu *rób, co chcesz* są konstrukcje *делай, что хочешь* oraz *делай, как хочешь*. Obie realizacje wykazują wysoką zgodność zarówno pod względem strukturalnym, jak i pragmatycznym. W materiale korpusowym odnotowano również przypadki, w których jednostka rosyjska *делай, что хочешь* tłumaczona była jako *możesz robić, co chcesz*. Choć liczba takich poświadczeń była niewielka, można przypuszczać, że wyrażenie to pełni zbliżone funkcje pragmatyczne i występuje w podobnych kontekstach komunikacyjnych. Z uwagi na ograniczoną liczbę konkordancji, zagadnienie to wymagałoby jednak osobnej, pogłębionej analizy. Należy podkreślić, że ogólna liczba

poświadczeń uzyskanych w obu korpusach była stosunkowo niska. Dotyczy to w szczególności jednostki *делай, что угодно*, która była najczęściej tłumaczona jako *zrób cokolwiek* – konstrukcja nieuwzględniona we wcześniejszych etapach analizy. Jednocześnie udało się potwierdzić typowe dla tej jednostki występowanie ograniczników (np. *делай что угодно, только не ходи туда* → *zrób cokolwiek, ale nie idź za nią na dół*). W związku z powyższym, jednostka ta nie została uznana za adekwatny translat frazy *rób, co chcesz*. Ostatecznie jako najbliższe ekwiwalenty pod względem zarówno struktury, jak i funkcji komunikacyjnej, można wskazać konstrukcje *делай, что хочешь* oraz *делай, как хочешь*.

Przejdźmy zatem do procesu ustalania odpowiednika czeskiego dla polskiej jednostki *rób, co chcesz*, co umożliwi finalne zestawienie trójjęzycznej triady przekładowej: polsko-rosyjsko-czeskiej. Podobnie jak w przypadku analizy rosyjskiej, procedura oparta zostanie na danych korpusowych, jednak ze względu na dostępność źródeł ograniczona zostanie do polsko-czeskiej części korpusu InterCorp v16ud. Zastosowano tę samą metodologię wyszukiwania, uwzględniając odpowiednie atrybuty zapytania, które pozwalają na precyzyjne odwzorowanie jednostki wyjściowej. Poniżej przedstawiono reprezentatywne przykłady konkordancji zarejestrowanych w korpusie, zawierające potencjalne odpowiedniki tłumaczeniowe w języku czeskim i określono ich frekwencję:

INTERCORP

☐ Kafka-Proces	- Klęcz albo czołgaj się na czworakach , rób , co chcesz , nic mnie to już nie obchodzi .	Kafka-Proces	* Klekni si nebo lez po všech čtyřech , dělej si co chceš . Já se o to starat nebudu . *
☐ Cortazar-nebe_peklo_ra	- Rób co chcesz , mnie to obojętne – mruknął Oliveira .	cortazar-nebe_peklo_ra	„Dělej , co chceš , mně je to jedno , “ řekl Oliveira .
☐ _SUBTITLES	- Rób , co chcesz .	_SUBTITLES	- Klidně , jak chceš .
☐ Ajtmatov-Popraviste	Idź i rób , co chcesz .	Ajtmatov-Popraviste	Běž a dělej , jak myslíš . *
☐ _SUBTITLES	Rób co chcesz .	_SUBTITLES	To je vaše věc , vy jste matka .

Ekwiwalent	Liczba poświadczeń danego ekwiwalentu	Procentowy udział w ogólnej liczbie poświadczeń
<i>Dělej si, co chceš</i>	196 poświadczeń	22%
<i>Dělej, co chceš</i>	57 poświadczeń	6,4%
<i>dělej, jak myslíš</i>	30 poświadczeń	3,3%
<i>Jak chceš</i>	18 poświadczeń	2%
<i>To je vaše věc</i>	2 poświadczenia	0,2%

W toku analizy polsko-czeskiej części korpusu InterCorp za potencjalnie najlepiej odpowiadający znaczeniowo i funkcjonalnie odpowiednik translandu uznano formę *Dělej si, co chceš*. Konstrukcja ta, w przeciwieństwie do polskiego i rosyjskiego odpowiednika, zawiera

częstkę *si*, którą można wstępnie interpretować jako formę *rób sobie, co chcesz*. Drugim najczęściej występującym ekwiwalentem w badanym materiale okazała się jednostka *Dělej, co chceš*, będąca dosłownym tłumaczeniem polskiej jednostki wyjściowej. W korpusie odnotowano ponadto nieliczne realizacje rzadziej używanych tłumaczeń, takich jak *Jak chceš*; *Dělej, jak myslíš* oraz *To je vaše věc*. Jak można zauważyć, udział wyodrębnionych jednostek w ogólnej liczbie poświadczeń w korpusie pozostaje relatywnie niski, podobnie jak w przypadku wyników uzyskanych w części polsko-rosyjskiej. W celu precyzyjnego ustalenia znaczeń wyodrębnionych jednostek wszystkie formy zostaną poddane dalszej analizie z wykorzystaniem czeskich korpusów jednojęzycznych: Araneum Bohemicum Maximum, InterCorp v16ud (wersja czeska) oraz korpusu mówionego ORAL v1. Semantyka jednostek została określona na podstawie analizy konkordancji, z uwzględnieniem rzeczywistych kontekstów użycia, a także opatrzone ją informacją o łącznej liczbie wystąpień:

EKWIWALENT	ZNACZENIE	PRZYKŁAD UŽYCIA
<i>Dělej si, co chceš</i>	1. Wyraz rzeczywistej obojętności. Wyrażenie sygnalizuje brak emocjonalnego zaangażowania nadawcy i pozostawienie rozmówcy pełnej swobody działania.	1. - Ty pokoje jsou sedmnáct let prázdné. I kdybych tam nechal udělat střelnici, mohlo by to mojí ženě být úplně jedno. Karen si odfrkla a zavrtěla hlavou. „<u>Dělej si, co chceš.</u> Zítra v poledne bude z mojí ošetrovny pryč. 2. Je to tvůj život. Nebudu ti bránit v rozletu. <u>Dělej si, co chceš.</u> Ale pokud budu mít dojem, že preferuješ a dáváš přednost obchodování přede mnou, je konec.
<i>Dělej, co chceš</i>		1. "Tak dost ,Morgrime. Ted' jdeme za tím naším člověkem a tam se teprve budeme rozhodovat, co dál. Já tohle chci mít co nejrychleji a nejtíšeji za sebou. Až to skončí, <u>dělej, co chceš.</u> ale ted' to, prosím, neztěžuj. 2. -O co tady jde, Stevene? Copak ty se mnou už nechceš být? - Jak jsi na to, sakra, přišla? -Tak si klidně <u>dělej, co chceš.</u> - Jak myslíš.
<i>Jak chceš</i>		1. Zkusím se rozhovor mezi vědmo u a Neem reprodukovat. Vědma: Posad' se. Neo :Raději postojím. V vědma: <u>Jak chceš.</u> Neo: Chtěl jsem sisednout. Vědma: Já vím. 2. Zamračeně na něj pohlédla, jako by v jeho slovech zaslechla něco, co možná ani nechtěl vyslovit – alespoň prozatím ne. Po chvíli zadumaného mlčení si však odevzdaně povzdechla. „<u>Jak chceš.</u> Vidím, že ti to nevymluví.
<i>Dělej, jak myslíš</i>		1. „Moc tomu nevěřím, “ pokrčila rameny, „možná bych ho už měla dát ve starým ohništi na pekáč. Nebo ho mám připravit jen nad

		plamenem na rožni? Co si o tom myslíš?“ -, <u>Dělej, jak myslíš.</u> 2. V tom se zmýlil, a je třeba dodat, že Leonid Iljič Gierka prorocky varoval: „<u>Dělej, jak myslíš,</u> ale ty i tvoje strana toho budete později litovat.
<i>To je vaše věc</i>	1. Zdystansowanie i przerzucenie odpowiedzialności. Wyrażenie sygnalizuje wycofanie się z rozmowy i brak dalszego zaangażowania. Nadawca przekazuje decyzję rozmówcy, dystansując się od jej skutków. 2. Pozorna zgoda połączona z dezaprobatą. Wyrażenie formalnie dopuszcza swobodę decyzji, lecz w podtekście zawiera niechęć, rozczarowanie lub brak akceptacji.	1. -Počítám za čtrnáct dní. -Cože? Chcete začít betonovat v období deštů? - Ano. -Myslíte to vážně? - Ano. -Dobře, <u>to je vaše věc.</u> A kdy začnete s přeložením hráze? 2. „Snad mne ještě pochopíte. Jen poslechněte mou radu.“ „Nevím, pane, zda jste člověkem, jehož radu bych mohl poslechnout.“ „Zda to uděláte, <u>to je vaše věc.</u> Ale doufám, že mou radu aspoň vyslechnete.

EKWIWALENT	FREKWENCJA
<i>Jak chceš</i>	10 521 poświadczeń
<i>Dělej si, co chceš</i>	1 195 poświadczeń
<i>Dělej, co chceš</i>	785 poświadczeń
<i>Dělej, jak myslíš</i>	474 poświadczenia
<i>To je vaše věc</i>	404 poświadczenia

Na podstawie przeprowadzonej analizy danych korpusowych można stwierdzić, że spośród potencjalnych czeskich odpowiedników tłumaczeniowych polskiego repoduktu *rób, co chcesz* największy stopień zgodności znaczeniowej i funkcjonalnej wykazuje jednostka *Dělej si, co chceš* (1 195 poświadczeń). Repodukt ten pełni dwie zasadnicze funkcje pragmatyczne: wyrażenie rzeczywistej obojętności oraz pozornej zgody z domniemaną dezaprobatą. Z kolei jednostka *Jak chceš*, mimo zdecydowanie najwyższej frekwencji w zasobach korpusowych (10 521 poświadczeń), została odrzucona jako bezpośredni translat ze względu na jej polifunkcyjność. W większości użyć występuje ona jako element szerszych struktur składniowych o odmiennym sensie (np. *Uděláme to, jak chceš ty; Říkej jim, jak chceš*;

Jak chceš, aby to bylo?). Pomimo to, szczegółowa analiza kontekstów użycia wskazuje na częściową zbieżność pragmatyczną z jednostką *Dělej si, co chceš*, przy czym *Jak chceš* pozbawione jest komponentu czasownikowego (*dělej si*), co zbliża je raczej do skróconego polskiego wariantu *jak chcesz*. Również *Dělej, co chceš* ujawnia podobną wielofunkcyjność i znaczeniowe pokrewieństwo, jednak zarejestrowana liczba wystąpień (785 poświadczeń) jest istotnie niższa w porównaniu z formą zawierającą zaimek zwrotny. Częściowym odpowiednikiem analizowanej jednostki może być czeskie wyrażenie *Dělej, jak myslíš* (474 poświadczenia), w którym dochodzi do przesunięcia semantycznego z modalności wolitywnej (*rób, co **chcesz***) na epistemiczną (dosł. *rób, jak **myślisz***), analogicznego do przesunięcia obecnego w polskim *Rób, jak uważasz*. W obu przypadkach decyzja jest prezentowana jako rezultat wewnętrznej oceny sytuacji przez rozmówcę, a nie jako efekt jego woli. Tego typu produkty występują zazwyczaj w kontekstach, w których nadawca odnosi się do wyboru między kilkoma możliwościami, a nie do pojedynczego działania, co wymaga od odbiorcy dokonania samodzielnej oceny. Niemniej jednak w analizowanych konkordancjach jednostka ta zachowuje znaczenie zbliżone do wcześniej omawianych translantów.

Ostatnia jednostka *To je vaše věc* (404 poświadczenia) pełni przede wszystkim funkcję przeniesienia odpowiedzialności na rozmówcę i sygnalizuje wycofanie się nadawcy z danej sytuacji. W większości przypadków wyraża rzeczywiste zdystansowanie, natomiast sporadycznie pojawia się w kontekstach pozornej zgody połączonej z krytyką. Z tego względu wykazuje większe podobieństwo do polskiego reproductu *to twoja sprawa* niż do *rób, co chcesz*, choć możliwa jest częściowa zbieżność funkcjonalna.

Końcowym etapem pozostaje pozakorpusowa analiza uzupełniająca, polegająca na weryfikacji wcześniej wyodrębnionych jednostek z wykorzystaniem innych źródeł. W dalszej części przedstawiono zestawienie jednostek, dla których udało się odnaleźć potwierdzenia lub dodatkowe informacje w wybranych źródłach, zgodnie z przyjętym poniżej schematem:

<i>Dělej si, co chceš</i>		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Glosbe	Fraza przetłumaczona jako „ <i>rób, co chcesz</i> ”.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Artykuł https://www.svobodauceni.cz/clan	Artykuł interpretuje frazę <i>Dělej si, co chceš</i> jako wypowiedź nacechowaną negatywnie, pełniącą funkcję	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

ek/delej-si-co-chces/ [data dostępu: 06.05.2025].	manipulacyjną i często sygnalizującą nadchodzący konflikt. Autor podkreśla, że zwrot ten bywa często używany przez rodziców wobec dzieci.	
F. Čermák, <i>Slovník české frazeologie a idiomatiky 4. Výrazy větné</i> , Nakladatelství LEDA, Praha 2009, s.203.	Wyrażenie <i>Dělej si, co chceš</i> interpretowane jako kolokwialne, które wyraża rezygnację i ukrytą niezgodę wobec decyzji rozmówcy, szczególnie mniej doświadczonego, a zarazem stanowi wyraz dystansowania się od jego wyboru.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>Dělej, co chceš</i>		
ŽRÓDLO	INFORMACJA	WNIOSEK
F. Čermák, <i>Slovník české frazeologie a idiomatiky 4. Výrazy větné</i> , Nakladatelství LEDA, Praha 2009, s. 203.	Wyrażenie <i>Dělej, co chceš</i> interpretowane jako kolokwialne, które wyraża rezygnację i ukrytą niezgodę wobec decyzji rozmówcy, szczególnie mniej doświadczonego, a zarazem stanowi wyraz dystansowania się od jego wyboru.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Artykuł https://prozeny.blesk.cz/clanek/pr-o-zeny-laska-a-vztahy/600034/5-beznych-frazi-ve-vztazich-ktere-muzi-a-zeny-chapou-jinak.html#1310130 [data dostępu: 06.05.2025].	W artykule fraza <i>Dělej, co chceš</i> jest przedstawiona jako forma manipulacji – pozorna zgoda, za którą kryje się dezaprobata. Kobiety używają jej, by uniknąć otwartej konfrontacji, jednocześnie wywierając presję emocjonalną.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>Jak chceš</i>		
ŽRÓDLO	INFORMACJA	WNIOSEK
Forum dyskusyjne https://www.emimino.cz/diskuse/jak-chces-482056/strankovani/2/ [data dostępu: 06.05.2025].	Fraza <i>Jak chceš</i> jest interpretowana, zwłaszcza w kontekście relacji partnerskich, jako wyraz emocjonalnej obojętności bądź braku zaangażowania w sytuację komunikacyjną. Niekiedy także jako przejaw niechęci do podjęcia jednoznacznej decyzji lub przyjęcia odpowiedzialności.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>Dělej, jak myslíš</i>		
ŽRÓDLO	INFORMACJA	WNIOSEK
Glosbe	Fraza przetłumaczona jako „zrobisz, co uważasz”.	Częściowe potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Ženský česko-český výkladový slovník https://www.fzu.cz/~sipr/documents/zensky_muzsky_slovník.html [data dostępu: 06.05.2025].	Niniejszy słownik humorystycznie interpretuje frazę <i>Dělej, jak myslíš</i> , która w rzeczywistości oznacza „budeš toho litovat” („będziesz tego żałować”). Mimo pozornej zgody, wyraża dezaprobatę i zapowiada negatywne konsekwencje.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
F. Čermák, <i>Slovník české frazeologie a idiomatiky 4. Výrazy větné</i> , Nakladatelství LEDA, Praha 2009, s. 203.	Wyrażenie <i>Dělej, jak myslíš</i> jest interpretowane jako kolokwialne i pojawia się najczęściej w sytuacji odrzucenia rady przez rozmówcę i wyraża zgodę na jego decyzję przy jednoczesnym zasygnalizowaniu odmiennego stanowiska nadawcy.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>To je vaše věc</i>		
ŽRÓDLO	INFORMACJA	WNIOSEK
Słownik Lingea https://slovniky.lingea.cz/polsko-cesky/v%C4%9Bc [data dostępu: 06.05.2025].	Choć artykuł hasłowy nie rejestruje analizowanego re produktu, odnotowuje parę <i>To je moje věc – To moja sprawa</i> , co może pośrednio potwierdzać wcześniej zaobserwowaną funkcję dystansującą – w tym przypadku o charakterze osobistym, a w przypadku jednostki <i>To je vaše věc</i> – odnoszącą się do rozmówcy.	Częściowe potwierdzenie poprzednich obserwacji w źródłach korpusowych.
F. Čermák, <i>Slovník české frazeologie a idiomatiky 4. Výrazy větné</i> , Nakladatelství LEDA, Praha 2009, s. 942.	Choć artykuł hasłowy nie rejestruje analizowanego re produktu, odnotowuje re produkt <i>To je moje věc</i> , który wyraża odmowę ingerencji i wyznacza granicę w relacji interpersonalnej. Jego funkcja dystansująca, odnoszącą się do siebie,	Częściowe potwierdzenie poprzednich obserwacji w źródłach korpusowych.

	może być potraktowana jako pośrednio zbieżna z funkcją jednostki <i>To je vaše věc</i> , choć nie pokrywa się z nią w sposób bezpośredni.	

Analiza przeprowadzona na podstawie źródeł pozakorpusowych potwierdza ustalenia wynikające z wcześniejszych badań korpusowych. Zgromadzony materiał wskazuje, że najtrafniejszymi odpowiednikami czeskimi dla polskiej jednostki *rób, co chceš* są konstrukcje *dělej si, co chceš, dělej, co chceš* oraz skrócona forma *jak chceš*. Warto podkreślić, że we wszystkich analizowanych źródłach podkreślano funkcjonowanie tych jednostek w kontekście wyrażania ukrytej dezaprobaty wobec decyzji rozmówcy, emocjonalnego dystansowania się, a niekiedy również potencjalnego użycia w funkcji manipulacyjnej. Co więcej, analiza ujawnia, że omawiane konstrukcje występują znacznie częściej w wypowiedziach kobiet oraz w relacjach nieformalnych, opartych na zażyłości rozmówców. Szczególnie często jednostki te pojawiają się w relacji rodzic – dziecko, gdzie nadawca zajmuje pozycję osoby z większym doświadczeniem. Zgodnie z wcześniejszymi hipotezami, fraza *to je vaše věc* nie może zostać zaliczona do funkcjonalnych ekwiwalentów analizowanej jednostki, gdyż pełni inne funkcje pragmatyczne – najczęściej sygnalizuje rzeczywiste wycofanie się z sytuacji decyzyjnej i brak ingerencji, a tylko sporadycznie przybiera formę pozornej zgody z jednoczesną dezaprobatą. Z kolei jednostka *jak chceš* może być uznana za częściowy ekwiwalent analizowanego translandu, ze względu na strukturę i znaczenie lepiej odpowiada polskiemu wyrażeniu *jak chceš*. Uwzględniając powyższe dane oraz częstość występowania poszczególnych jednostek w poświadczeniach korpusowych, możliwe jest sformułowanie propozycji par przekładowych uporządkowanych według stopnia ich odpowiedniości do jednostki wyjściowej. Najwyższy poziom zgodności znaczeniowej i funkcjonalnej wykazuje para:

1. *rób, co chceš* -> *dělej si, co chceš*; a następnie
2. *rób, co chceš* -> *dělej, co chceš*.

Ostatnim etapem analizy jest wykonanie przekładu odwrotnego z wykorzystaniem czesko-polskiej wersji korpusu InterCorp v16ud, w którym ustalone wcześniej jednostki czeskie traktowane są jako transland. Poniżej zaprezentowano zestawienie odpowiadających im jednostek polskich wraz z informacją o ich częstości występowania w korpusie:

TRANSLAND CZESKI	TRANSLAT POLSKI	FREKWENCJA
<i>dělej si, co chceš</i>	<i>rób, co chcesz</i>	104 poświadczenia (30%)
	<i>możesz robić, co chcesz</i>	21 poświadczeń (6,1%)
	<i>rób, jak chcesz</i>	19 poświadczeń (5,6%)
<i>dělej, co chceš</i>	<i>rób, co chcesz</i>	76 poświadczeń (38%)
	<i>możesz robić, co chcesz</i>	7 poświadczeń (3,5%)

Analiza przekładu odwrotnego z języka czeskiego potwierdziła, że spośród dostępnych realizacji to konstrukcje *dělej si, co chceš* oraz *dělej, co chceš* najczęściej funkcjonują jako odpowiedniki polskiego re produktu *rób, co chcesz*. Ich wysoką frekwencję oraz zgodność funkcjonalną i strukturalną potwierdzają zarówno dane korpusowe, jak i wcześniejsze obserwacje uzyskane w toku analizy materiałów pozakorpusowych. W uzyskanych konkordancjach zaobserwowano również jednostkowe przypadki zastosowania wyrażenia *możesz robić, co chcesz*, analogiczne do tych odnotowanych wcześniej w analizie przekładu rosyjsko-polskiego. W korpusie odnotowano również pojedyncze przypadki przekładu jednostki *dělej si, co chceš* jako *rób, jak chcesz*, jednak ze względu na ich marginalną frekwencję nie miały one istotnego wpływu na finalne ustanowienie adekwatnych par przekładowych. W świetle powyższych ustaleń, za najtrafniejsze ekwiwalenty należy uznać konstrukcje *dělej si, co chceš* oraz *dělej, co chceš*, których użycie znajduje potwierdzenie zarówno w danych korpusowych, jak i źródłach pozakorpusowych.

Na podstawie przeprowadzonej analizy metodą korpusową oraz weryfikacji danych w źródłach pozakorpusowych można wskazać, że najtrafniejszymi odpowiednikami tłumaczeniowymi polskiego re produktu *rób, co chcesz* są rosyjskie konstrukcje *делай, что хочешь* i *делай, как хочешь* oraz czeskie wyrażenia *dělej si, co chceš* i *dělej, co chceš*. Wszystkie wymienione jednostki wykazują wysoką zgodność pod względem strukturalnym i funkcjonalnym, co potwierdzają zarówno dane z korpusów równoległych, jak i jednojęzycznych. Przeprowadzona analiza kontekstów użycia pozwala również określić podstawowe funkcje pragmatyczne analizowanego translandu i uzyskanych translatów. Analizowana fraza najczęściej pełni funkcję pozornej zgody, nasyconej elementem ukrytej dezaprobaty, rozczarowania lub emocjonalnego dystansu. Znacznie rzadziej występuje w kontekstach, w których sygnalizuje rzeczywistą obojętność i brak zaangażowania nadawcy. Jednostka ta bywa używana szczególnie w relacjach opartych na bliskości i asymetrii

doświadczeń – np. między rodzicami a dziećmi, partnerami lub osobami, spośród których jedna zajmuje wyraźnie dominującą pozycję. Zgromadzony materiał wskazuje również na częstsze występowanie omawianych konstrukcji w wypowiedziach kobiet, co może świadczyć o ich zachowaniach w sytuacjach nacechowanych emocjonalnie. Analiza potwierdza zatem, że mimo pozornej prostoty, reprodukt *rób, co chcesz* funkcjonuje jako nośny środek wyrażania postaw interpersonalnych, mocno zakorzeniony w relacjach społecznych.

POLSKI	ROSYJSKI	CZESKI
1. <i>rób, co chcesz</i>	1. <i>делай, что хочешь</i>	1. <i>dělej si, co chceš</i>
	2. <i>делай, как хочешь</i>	2. <i>dělej, co chceš</i>

Fatyzmy

W danym miejscu podjęta zostanie analogiczna próba ustalenia rosyjskich i czeskich ekwiwalentów przekładowych dla jednego z bardziej niejednoznacznych fatyzmów, pełniącego funkcję nie tylko inicjowania, lecz także podtrzymywania wypowiedzi, jakim jest wyrażenie *szczerze mówiąc*. W WSJP PAN autorzy przytaczają następującą definicję: ‘mówiący chce podkreślić, że to, co mówi, jest zgodne z tym, co sądzi, ale wahał się, czy powinien to mówić’. Nie ulega wątpliwości, że powyższa interpretacja odpowiada potocznemu rozumieniu wyrażenia przez rodzimego użytkownika języka polskiego. Z drugiej jednak strony pozostaje ona stosunkowo ogólna, ponieważ nie wyszczególnia najważniejszych funkcji znaczeniowych ani typowych kontekstów użycia. Na podstawie uprzednio przeprowadzonych obserwacji i analiz uznano, że spośród wariantów takich jak *szczerze powiedziawszy*; *tak szczerze mówiąc* czy skrócone *szczerze*, najbardziej rozpowszechniona i funkcjonalnie samodzielna jest forma *szczerze mówiąc*, która w dalszej analizie traktowana będzie jako inwariantywna.

Weryfikacja frekwencji w trzech wspomnianych już korpusach, tj. NKJP, Araneum Polonicum Maius oraz polskiej części korpusu InterCorp v16ud – wykazała łącznie 13 718 poświadczeń użycia analizowanego wyrażenia (NKJP – 1795, APM – 6232, InterCorp – 5691). Należy zaznaczyć, iż w języku mówionym występuje również forma odwrócona, tj. *mówiąc szczerze*, która – choć znaczeniowo tożsama – cechuje się niższą frekwencją. Z tego względu dalsza analiza koncentrować się będzie na bardziej rozpowszechnionej formie podstawowej. Tak znaczna liczba konkordancji dla fatyzmu *szczerze mówiąc* umożliwiła przeprowadzenie szczegółowej analizy głównych funkcji pragmatycznych i znaczeniowych tej jednostki, której wyniki przedstawiono poniżej:

FATYZM	ZNACZENIE	PRZYKŁAD UŻYCIA
1. <i>szczerze mówiąc</i>	Zaznaczenie zgodności wypowiedzi z rzeczywistym przekonaniem mówiącego. Mówiący podkreśla, że wypowiedź odzwierciedla jego osobiste zdanie lub pogląd, który uznaje za prawdziwy, a jej celem jest wyrażenie szczerości i otwartości wobec odbiorcy.	- <i>Nie. Daję panu słowo, że o tym nie myślałem.</i> - <i>Aż słowo! - roześmiał się Maliński. - Ja, <u>szczerze mówiąc</u>, już bym nie próbował. Ale to nie znaczy, żeby jeszcze jakieś próby były całkiem pozbawione sensu.</i>
2. <i>szczerze mówiąc</i>	Przyznanie się do wahania lub wewnętrznego oporu. Wyrażenie sygnalizuje trudność w wypowiedzeniu danej treści. Mówiący daje do zrozumienia, że waha się bądź powstrzymuje od jej ujawnienia	- <i>Panie przewodniczący, to jest, <u>szczerze mówiąc</u>, wolałbym... bo ja się czuję niewygodnie z tą sprawą, no ale skoro takie pytanie padło, to pewnie muszę na to odpowiedzieć.</i>
3. <i>szczerze mówiąc</i>	Zapowiedź sprzeciwu, opinii krytycznej lub niepopularnej. Jednostka pełni funkcję amortyzującą – przygotowuje odbiorcę na wypowiedź, która może być niezgodna z jego oczekiwaniami, wzbudzać sprzeciw lub zawierać treść trudną do przyjęcia.	- <i>Problem w tym, że osoby z takim wyrokiem nie mają go gdzie odpracować. Prezes sądu, Jadwiga Noworyta stwierdza, że dużo do zrobienia w tej sprawie mają samorządy. - <u>Szczerze mówiąc</u> dziwi mnie brak zaangażowania ze strony samorządów, bo przecież prac jest bardzo dużo do zrobienia w każdej gminie i powiecie.</i>
4. <i>szczerze mówiąc</i>	Przyznanie się do winy lub błędu. Jednostka służy do zasygnalizowania autorefleksji i przyjęcia odpowiedzialności za niepowodzenie, pomyłkę lub niewłaściwe postępowanie. Ma charakter samokrytyczny.	- <i><u>Szczerze mówiąc</u> to moja wina, bo to ja powiedziałem, że chcę strzelić 1000 i teraz gdy przekroczyłem 900, już pytają mnie o więcej. Tymczasem nie chcę teraz myśleć długoterminowo. Wolę się cieszyć chwilą.</i>

5. <i>szczerze mówiąc</i>	Podkreślenie zgodności opisu z rzeczywistością w wypowiedzi neutralnej lub sprawozdawczej. Mówiący nie ocenia i nie wyraża emocji, lecz zaznacza, że przedstawiany stan rzeczy jest zgodny z faktami, jak je postrzega.	- <i>Nic nie sędzę - powiedział. - Chodzi mi o tych dwóch mężczyzn. Znała ich pani?</i> - <i>Przysiedli się do nas, przecież pan wie, jak to bywa... dwie samotne młode kobiety... Niebrzydkie. - Uśmiechnęła się znowu. – <u>Szczerze mówiąc</u>, nie miałyśmy nic przeciwko temu. Obaj byli przystojni i grzeczni.</i>
6. <i>szczerze mówiąc</i>	Przyznanie się do niewiedzy lub braku stanowiska. Mówiący sygnalizuje brak wiedzy, niepewność lub niemożność sformułowania jednoznacznej opinii. Może mieć charakter samoobronny lub dystansujący.	<i>Rzeczywiście, panie pośle, ma pan rację, jest tam zapis, że na wniosek 1/4 ustawowego składu sejmiku i nie ma zapisu, który by dawał jakkolwiek płynność <u>Szczerze mówiąc</u> nie wiem, co można dzisiaj zrobić, żeby to zmienić.</i>
7. <i>szczerze mówiąc</i>	Przyznanie się do zaskoczenia. Mówiący sygnalizuje, że odbiór danej sytuacji, informacji lub zdarzenia był sprzeczny z przewidywaniami mówiącego. Może pełnić funkcję wyrażającą autentyczne zdziwienie, konsternację lub niedowierzanie – zarówno pozytywne, jak i negatywne.	- <i><u>Szczerze mówiąc</u> nie spodziewałem się aż tak dobrej imprezy. Bardzo się cieszę, że mogłem pomóc chłopakom w organizacji ich urodzin. Dzięki tak udanym wydarzeniom wszyscy jeszcze bardziej zachęcamy się do takich przedsięwzięć.</i>

Analiza materiału korpusowego ujawniła wieloznaczność oraz wielość funkcji pragmatycznych wyrażenia *szczerze mówiąc*, co potwierdza jego użycie w różnych kontekstach komunikacyjnych. Już na tym etapie badań widoczna jest złożoność analizowanego fatyzmu nie tylko z perspektywy translacyjnej, lecz także jednojęzycznej. Jednostki tego typu, często marginalizowane jako elementy „łatwe” do przełożenia, w rzeczywistości mogą ujawniać wysoki stopień wieloznaczności i niestabilności funkcjonalnej. Zgromadzone konkordancje dowodzą, że znaczenia tej jednostki są silnie zróżnicowane, a niektóre z nich mogą się wzajemnie przenikać. Tym samym jej analiza w ujęciu jednojęzycznym okazała się zadaniem

wymagającym, sugerującym, że fatyzmy mogą w praktyce charakteryzować się większym stopniem złożoności i wieloznaczności, niż się powszechnie zakłada. Na podstawie przeprowadzonej analizy wyodrębniono siedem podstawowych funkcji wyrażenia, spośród których najczęściej realizowane były:

- zaznaczenie zgodności wypowiedzi z osobistym przekonaniem mówiącego,
- wyrażenie zaskoczenia oraz
- zastosowanie w wypowiedziach sprawozdawczych.

Tak szerokie spektrum funkcji utrudnia proces przekładu – optymalnym rozwiązaniem byłoby odnalezienie jednostki docelowej zdolnej do realizowania wszystkich wskazanych znaczeń. Zgodnie z procedurą zastosowaną uprzednio dla jednostki emotywnej, kolejnym krokiem pozostaje wprowadzenie wyrażenia *szczerze mówiąc* do:

1. wyszukiwarki korpusu równoległego polsko-rosyjskiego InterCorp v16ud, z zachowaniem tych samych atrybutów;
2. polsko-rosyjskiej części korpusu równoległego NKJR, z wykorzystaniem opcji „Поиск точных форм”.

Poniżej zaprezentowano zestawienie tłumaczeń uzyskanych w obu źródłach, uwzględniające odpowiedniki najbardziej zbliżone znaczeniowo wraz z informacją o częstotliwości ich występowania jako realizacji par przekładowych:

INTERCORP

☐ ishiguro-neopoustej_me	- Szczerze mówiąc - oznajmiła - od samego początku wiedziała m , że to głupota .	ishiguro-neopoustej_me	– Честно говоря , я с самого начала знала , что это глупость .
☐ ishiguro-neopoustej_me	Szczerze mówiąc , nie wiem nawet , czy to ona pierwsza wspomniała o zwierzętach .	ishiguro-neopoustej_me	Если честно , я даже не могу поручиться , что она первая упомянула его рисунки .
☐ Kundera-Nesnesit_Jehko	Szczerze mówiąc , nie umiała by nawet go opisać , chyba w ogóle nie zwróciła uwagi na to , jak wyglądał nago .	Kundera-Nesnesit_Jehko	Откровенно говоря , она бы даже не смогла его описать , она , пожалуй , и не заметила , как выглядел он голый .
☐ larsson-hnizdo	Szczerze mówiąc , nie istnieje nawet dokumentacja potwierdzająca , że to pan Zalachenko pobił matkę Lisbeth Salander .	larsson-hnizdo	По правде говоря , отсутствуют даже документы , подтверждающие , что мать Саландер избил именно господин Залаченко .

Ekwiwalent	Liczba poświadczeń danego ekwiwalentu	Procentowy udział w ogólnej liczbie poświadczeń
<i>честно говоря</i>	606	23%
<i>если честно</i>	459	17,4%
<i>откровенно говоря</i>	91	3,4%
<i>по правде говоря</i>	75	2,8%

польский: - Drobną różnicą zda? na temat metod działania w stosunkach służbowych - wyjaśnił Marek uprzejmie. - <i>Szczerze mówiąc</i> nic z tego nie rozumiem. 🗨️ ↔️	русский: - Небольшое расхождение взглядов относительно служебных контактов, – любезно объяснил Марек. – Честно говоря, я в этом ничего не понимаю. 🗨️ ↔️

1. Andrzej Sapkowski. Lux Perpetua (2) (2006) | Анджей Сапковский. Свет вечный (2) (В. Фляк, 2009) ⓘ

польский: <i>Szczerze mówiąc</i>, wcale ci się nie dziwiłem. 🗨️ ↔️	русский: Откровенно говоря, я вовсе тебе не удивился. 🗨️ ↔️

Ekwiwalent	Liczba poświadczeń danego ekwiwalentu	Procentowy udział w ogólnej liczbie poświadczeń
<i>честно говоря</i>	4 poświadczenia	30,7%
<i>откровенно говоря</i>	4 poświadczenia	30,7%

Przeprowadzone badanie w obu korpusach równoległych umożliwiło wyodrębnienie kilku potencjalnych odpowiedników tłumaczeniowych wyrażenia *szczerze mówiąc*. Najczęściej rejestrowane były jednostki *честно говоря* (tłumaczenie dosłowne), *если честно* oraz *откровенно говоря*, które wykazują największą zgodność w stosunku do jednostki wyjściowej. W przeciwieństwie do nich forma *по правде говоря* pojawiała się sporadycznie i zdaje się być bliższa innemu polskiemu translandowi – *prawdę mówiąc*. Ogólna liczba analizowanych par przekładowych pozyskanych z obu korpusów okazała się zadowalająca w porównaniu z wynikami poprzedniej analizy — łącznie zidentyfikowano 2619 wystąpień, z czego zdecydowaną większość dostarczył korpus InterCorp. Warto zwrócić uwagę na brak rejestracji formy *если честно* w korpusie NKJR najprawdopodobniej wynika z jego ograniczonych polsko-rosyjskich zbiorów, co istotnie zawęża możliwości obserwacji i porównania częstości występowania tej jednostki. Podkreślenia wymaga również fakt, iż w materiale korpusowym nie zaobserwowano ani wymienności komponentów, ani występowania form odwróconych, co skłania do wniosku, że analizowane jednostki funkcjonują jako reproduktory leksykalnie zamknięte, w których poszczególne elementy zajmują stałe i niepodlegające modyfikacji miejsce w strukturze.

Wobec powyższego, wszystkie wyodrębnione jednostki poddane zostaną dalszej analizie w wybranych rosyjskich korpusach jednojęzycznych: Araneum Russicum Maius, rosyjskiej części InterCorp v16ud oraz NKJR. Celem jest ustalenie możliwie pełnego spektrum znaczeń i funkcji każdej z jednostek w autentycznych kontekstach użycia. Poniżej

przedstawiono szczegółowe zestawienie zidentyfikowanych znaczeń wraz z przykładami ilustrującymi ich funkcjonowanie w rzeczywistych danych językowych oraz informacją o częstotliwości ich występowania w analizowanych zbiorach:

EKWIWALENT	ZNACZENIE	ПРИКЛАД УЖИЦА
1. <i>честно говоря</i>	1. Zaznaczenie zgodności wypowiedzi z rzeczywistym przekonaniem mówiącego. Mówiący podkreśla, że wypowiedź odzwierciedla jego osobiste zdanie lub pogląd, który uznaje za prawdziwy, a jej celem jest wyrażenie szczerości i otwartości wobec odbiorcy. .	- Когда говорят, что родовая боль забывается – это всё враньё. Я родила 3 недели назад. И не думаю, что со временем забуду, что это такое. <u>Честно говоря</u> , я бы памятник при жизни поставила тем женщинам, которые решились родить второго, третьего и т.д.
2. <i>честно говоря</i>	Przyznanie się do wahania lub wewnętrznego oporu. Wyrażenie sygnalizuje trudność w wypowiedzeniu danej treści. Mówiący daje do zrozumienia, że waha się bądź powstrzymuje od jej ujawnienia	- Знаешь, <u>честно говоря</u> мне стыдно признаваться, но я даже в какой-то степени ожидал поездки туда.
3. <i>честно говоря</i>	Zapowiedź sprzeciwu, opinii krytycznej lub niepopularnej. Jednostka pełni funkcję amortyzującą – przygotowuje odbiorcę na wypowiedź, która może być niezgodna z jego oczekiwaniami, wzbudzać sprzeciw lub zawierać treść trudną do przyjęcia.	Вот цитаты студентов из ДонНУ": филфак хорош в плане того, что там учится очень много разносторонне развитых людей, творческих личностей... но, <u>честно говоря</u> , журналистике там не научат. большинство быстро разочаровывается.
4. <i>честно говоря</i>	Podkreślenie zgodności opisu z rzeczywistością w wypowiedzi neutralnej lub sprawozdawczej. Mówiący nie ocenia i nie wyraża emocji, lecz zaznacza, że przedstawiany stan rzeczy jest zgodny z faktami, jak je postrzega.	- Как в США оценивают тюремное заключение Тимошенко? - <u>Честно говоря</u> , в Соединенных Штатах общество не особо много внимания уделяет

		Украине. Нужно быть реалистом.
5. <i>честно говоря</i>	Przyznanie się do niewiedzy lub braku stanowiska. Mówiący sygnalizuje brak wiedzy, niepewność lub niemożność sformułowania jednoznacznej opinii. Może mieć charakter samoobronny lub dystansujący.	- Но, если не ошибаюсь, во время приступа эпилепсии язык прикушен, как же он западёт? [...] А во время приступа кислород в любом случае не поступает. Может я не прав, <u>честно говоря</u> , не знаю.
6. <i>честно говоря</i>	Przyznanie się do zaskoczenia. Mówiący sygnalizuje, że odbiór danej sytuacji, informacji lub zdarzenia był sprzeczny z przewidywaniami mówiącego. Może pełnić funkcję wyrażającą autentyczne zdziwienie, konsternację lub niedowierzanie – zarówno pozytywne, jak i negatywne.	- Поздравляю всех с окончанием 2 -ой лиги на Нохwin! <u>Честно говоря</u> не ожидал, что для меня сложится все так удачно. Хорошо, что первый тур был бесплатный.
7. <i>честно говоря</i>	Przyznanie się do winy lub błędu. Jednostka służy do zasygnalizowania autorefleksji i przyjęcia odpowiedzialności za niepowodzenie, pomyłkę lub niewłaściwe postępowanie. Ma charakter samokrytyczny.	-Мне непонятно, почему вы не сдались, хоть на время, зачем оскорбили его? - <u>Честно говоря</u> , это была огромная ошибка. - Ошибка? Мы могли бы вывесить уведомления о выселении прямо сейчас...

EKWIWALENT	ZNACZENIE	PRZYKŁAD UŻYCIA
1. <i>если честно</i>	Zaznaczenie zgodności wypowiedzi z rzeczywistym przekonaniem mówiącego. Mówiący podkreśla, że wypowiedź odzwierciedla jego osobiste zdanie lub pogląd, który uznaje za prawdziwy, a jej celem jest wyrażenie szczerości i otwartości wobec odbiorcy.	- Я не знала, что в городе собираются поставить памятник. И <u>если честно</u> , мне все равно», — говорит лаборант Таусия.

	.	
2. <i>если честно</i>	Przyznanie się do wahania lub wewnętrznego oporu. Wyrażenie sygnalizuje trudność w wypowiedzeniu danej treści. Mówiący daje do zrozumienia, że waha się bądź powstrzymuje od jej ujawnienia	- <i>Что случилось ? - Я в порядке. - Что произошло? Я не очень хочу говорить об этом , <u>если честно</u>.</i>
3. <i>если честно</i>	Zapowiedź sprzeciwu, opinii krytycznej lub niepopularnej. Jednostka pełni funkcję amortyzującą – przygotowuje odbiorcę na wypowiedź, która może być niezgodna z jego oczekiwaniami, wzbudzać sprzeciw lub zawierać treść trudną do przyjęcia.	- <i>В мире было только две страны, поддерживавших полный спектр фундаментальных исследований – СССР и США. И без первой, вторая в этом смысле хиреет. <u>Если честно</u>, то твои рецепты по администрированию науки – это детство.</i>
4. <i>если честно</i>	Podkreślenie zgodności opisu z rzeczywistością w wypowiedzi neutralnej lub sprawozdawczej. Mówiący nie ocenia i nie wyraża emocji, lecz zaznacza, że przedstawiany stan rzeczy jest zgodny z faktami, jak je postrzega.	- <i>Теперь давайте по прибыли. <u>Если честно</u>, то в среднем такой сайт приносит порядка 450 - 500 рублей в месяц.</i>
5. <i>если честно</i>	Przyznanie się do niewiedzy lub braku stanowiska. Mówiący sygnalizuje brak wiedzy, niepewność lub niemożność sformułowania jednoznacznej opinii. Może mieć charakter samoobronny lub dystansujący.	- <i>Что изменилось с тех пор в Верхотурье? — <u>Если честно</u>, не знаю. - А вы уже разбирались в этом вопросе? — Что-то, безусловно, сделано.</i>
6. <i>если честно</i>	Przyznanie się do zaskoczenia. Mówiący sygnalizuje, że odbiór danej sytuacji, informacji lub zdarzenia był sprzeczny z przewidywaniami mówiącego. Może pełnić funkcję wyrażającą autentyczne zdziwienie, konsternację lub niedowierzanie – zarówno pozytywne, jak i negatywne.	- <i>Всех , кто сегодня выступал, поздравляю. <u>Если честно</u>, думал, что медалей будет значительно меньше. Вы молодцы. Все, кто отстрелялся , свободны.</i>

7. <i>если честно</i>	Przyznanie się do winy lub błędu. Jednostka służy do zasygnalizowania autorefleksji i przyjęcia odpowiedzialności za niepowodzenie, pomyłkę lub niewłaściwe postępowanie. Ma charakter samokrytyczny.	- Но мне кажется, <u>если честно</u> , я слегка не рассчитал свои силы. Вовсе нет.

EKWIWALENT	ZNACZENIE	ПРИКЛАД УЖИТИЯ
1. <i>откровенно говоря</i>	Zaznaczenie zgodności wypowiedzi z rzeczywistym przekonaniem mówiącego. Mówiący podkreśla, że wypowiedź odzwierciedla jego osobiste zdanie lub pogląd, który uznaje za prawdziwy, a jej celem jest wyrażenie szczerości i otwartości wobec odbiorcy.	- Мой друг Слава Хаит выбрал другой путь — остался на свободе. [...] дома его никто не ждет. И, <u>откровенно говоря</u> , иногда я ему завидую.
2. <i>откровенно говоря</i>	Przyznanie się do wahania lub wewnętrznego oporu. Wyrażenie sygnalizuje trudność w wypowiedzeniu danej treści. Mówiący daje do zrozumienia, że waha się bądź powstrzymuje od jej ujawnienia	-Послушайте, Алекс, я ... я действительно рад, что вы вернулись. <u>Откровенно говоря</u> , я был ... был немного обеспокоен после вашего ухода на прошлой неделе.
3. <i>откровенно говоря</i>	Zapowiedź sprzeciwu, opinii krytycznej lub niepopularnej. Jednostka pełni funkcję amortyzującą – przygotowuje odbiorcę na wypowiedź, która może być niezgodna z jego oczekiwaniami, wzbudzać sprzeciw lub zawierać treść trudną do przyjęcia.	- Для этого нужно хорошенько растереться полотенцем. И тут возникают вопросы: как можно растираться тем, что <u>откровенно говоря</u> , напоминает наждачную бумагу, а не полотенце.
4. <i>откровенно говоря</i>	Podkreślenie zgodności opisu z rzeczywistością w wypowiedzi neutralnej lub sprawozdawczej. Mówiący nie ocenia i nie wyraża emocji, lecz zaznacza, że	Пансион скучных тётток. Учреждение это никогда не пустовало и вечно скучало. <u>Откровенно говоря</u> , жители городка побаивались туда

	przedstawiany stan rzeczy jest zgodny z faktami, jak je postrzega.	заглядывать, и даже когда ходили за грибами в Лес смыслов.
5.откровенно говоря	Przyznanie się do niewiedzy lub braku stanowiska. Mówiący sygnalizuje brak wiedzy, niepewność lub niemożność sformułowania jednoznacznej opinii. Może mieć charakter samoobronny lub dystansujący.	- Ну и сходи. Что за проблема ? - Откровенно говоря , я не знаю, с чего начать, что делать... - План писать! - вдруг взорвался он.
6.откровенно говоря	Przyznanie się do zaskoczenia. Mówiący sygnalizuje, że odbiór danej sytuacji, informacji lub zdarzenia był sprzeczny z przewidywaniami mówiącego. Może pełnić funkcję wyrażającą autentyczne zdziwienie, konsternację lub niedowierzanie – zarówno pozytywne, jak i negatywne.	Я, откровенно говоря , был потрясен информацией о том, что в разные времена [...] Россия направила в наши союзные республики более 25-ти миллионов россиян - специалистов высокого класса.
7.откровенно говоря	Przyznanie się do winy lub błędu. Jednostka służy do zasygnalizowania autorefleksji i przyjęcia odpowiedzialności za niepowodzenie, pomyłkę lub niewłaściwe postępowanie. Ma charakter samokrytyczny.	- Я не хочу сказать, что вы пытались себя выгородить, скорее наоборот! Откровенно говоря я не знаю, почему не уволила вас сразу же.

EKWIWALENT	ZNACZENIE	PRZYKŁAD UŻYCIA
1. по правде говоря	Zaznaczenie zgodności wypowiedzi z rzeczywistym przekonaniem mówiącego. Mówiący podkreśla, że wypowiedź odzwierciedla jego osobiste zdanie lub pogląd, który uznaje za prawdziwy, a jej celem jest wyrażenie szczerości i otwartości wobec odbiorcy.	- А меня, по правде говоря , — сказал тогда Катул, — всегда особенно восхищает, когда при всей несхожести вашего красноречия каждый из вас говорит так, что [...].
2. по правде говоря	Przyznanie się do wahania lub wewnętrznego oporu. Wyrażenie sygnalizuje trudność w wypowiedzeniu	- Я люблю тебя. И мне жаль... Но я лгал, когда я дал тебе кольцо...и сказал, что

	danej treści. Mówiący daje do zrozumienia, że waha się bądź powstrzymuje od jej ujawnienia	<i>это ничего не значит. По правде говоря, я... Я хотел спросить вас жениться меня прямо тогда и там.</i>
<i>3. по правде говоря</i>	Zapowiedź sprzeciwu, opinii krytycznej lub niepopularnej. Jednostka pełni funkcję amortyzującą – przygotowuje odbiorcę na wypowiedź, która może być niezgodna z jego oczekiwaniami, wzbudzać sprzeciw lub zawierać treść trudną do przyjęcia.	<i>- По правде говоря, ты можешь иногда слишком быстро давать советы. - Ну, я всегда готова помочь, если ты об этом, то да.</i>
<i>4. по правде говоря</i>	Podkreślenie zgodności opisu z rzeczywistością w wypowiedzi neutralnej lub sprawozdawczej. Mówiący nie ocenia i nie wyraża emocji, lecz zaznacza, że przedstawiany stan rzeczy jest zgodny z faktami, jak je postrzega.	<i>- По правде говоря, Аньес не любила ни чайный сервиз, ни кресло- качалку, ни рояль. Не то, чтобы эти вещи были безвкусны , но в них во всех было нечто эксцентричное [...].</i>
<i>5. по правде говоря</i>	Przyznanie się do niewiedzy lub braku stanowiska. Mówiący sygnalizuje brak wiedzy, niepewność lub niemożność sformułowania jednoznacznej opinii. Może mieć charakter samoobronny lub dystansujący.	<i>- Вы общаетесь с ним? - Нет. Уже нет. По правде говоря, я не знаю, где он теперь. Отец обещал взять меня на матч в субботу</i>
<i>6. по правде говоря</i>	Przyznanie się do zaskoczenia. Mówiący sygnalizuje, że odbiór danej sytuacji, informacji lub zdarzenia był sprzeczny z przewidywaniami mówiącego. Może pełnić funkcję wyrażającą autentyczne zdziwienie, konsternację lub niedowierzanie – zarówno pozytywne, jak i negatywne.	<i>[...] я чувствовал сладкий, тягучий запах хереса, который она пила, и не успел я опомниться, как ее губы уже прильнули к моим. По правде говоря, я вовсе этого не ожидал – я повернулся, просто чтобы посмотреть на ее лицо.</i>
<i>7. по правде говоря</i>	Przyznanie się do winy lub błędu. Jednostka służy do zasygnalizowania autorefleksji i przyjęcia odpowiedzialności za niepowodzenie, pomyłkę lub	<i>- Как только ты сюда вошел впервые, было понятно, что с тобой что-то творится. По правде говоря, мне немного</i>

	niewłaściwe postępowanie. Ma charakter samokrytyczny.	<i>стыдно было, что я тобой воспользовалась.</i>

EKWIWALENT	FREKWENCJA
<i>честно говоря</i>	18 848 poświadczeń
<i>если честно</i>	12 200 poświadczeń
<i>откровенно говоря</i>	3276 poświadczeń
<i>по правде говоря</i>	3305 poświadczeń

Przeprowadzona analiza potwierdziła, że wszystkie wyodrębnione rosyjskie jednostki potencjalnie odpowiadające polskiemu wyrażeniu *szczerze mówiąc* pełnią zbliżone funkcje komunikacyjne. Należy również podkreślić, iż w badanym materiale wszystkie z zaprezentowanych fatyzmów występowały na początku lub w środku wypowiedzi, pełniąc funkcję środka podtrzymującego komunikację.

Największą zgodność znaczeniową oraz najwyższą frekwencję odnotowano dla jednostki *честно говоря* (18 848 poświadczeń). Równie wysoką częstością charakteryzuje się *если честно* (12 200 poświadczeń), które może być funkcjonalnie bliższe bardziej rozbudowanej formie polskiej, takiej jak *jeśli mam być szczerzy/szczera*. Znacznie mniejszą liczbę poświadczeń posiadają jednostki *откровенно говоря* (3276) oraz *по правде говоря* (3305), jednak i one realizują analogiczne funkcje pragmatyczne, co czyni je potencjalnie prawidłowymi odpowiednikami dla form typu *otwarcie mówiąc* lub *prawdę mówiąc*. Istotnym wnioskiem płynącym z analizy jest fakt, że wszystkie jednostki ujawniają wysoką zgodność funkcjonalną i występują w porównywalnych sytuacjach komunikacyjnych, co uzasadnia ich użycie jako ekwiwalentów rosyjskich. Można zatem z dużym prawdopodobieństwem uznać je za trafne odpowiedniki przekładowe, przy czym szczególną uwagę należy zwrócić na ich frekwencję w użyciu – za najbliższe polskiemu wyrażeniu uznano *честно говоря* oraz *если честно*. Końcowym etapem przed ostatecznym ustaleniem adekwatnych par przekładowych pozostaje weryfikacja funkcjonowania tych jednostek w źródłach pozakorpusowych. W dalszej części przedstawiono wyniki analizy materiału pozakorpusowego, koncentrując się na sposobach interpretacji wybranych jednostek w poszczególnych źródłach oraz na przypisywanych im funkcjach pragmatycznych:

честно говоря		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Rosyjski słownik jednojęzyczny Кузнецов С.А., <i>Большой толковый словарь русского языка</i> , Издательство Норинт, Санкт-Петербург 2000, с. 1476.	W haśle <i>честно</i> odnotowano jednostkę <i>честно говоря</i> , interpretowaną jako wtrącenie i pełniącą przy tym funkcję sygnalizowania szczerości i bezpośredniości w wyrażaniu stanowiska nadawcy.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Internetowy słownik frazeologiczny phraseologiya.academic.ru/1127/честно_говоря [data dostępu: 10.05.2025].	Jednostka została scharakteryzowana jako wyrażenie potoczne, służące wyrażeniu opinii w sposób otwarty, bezpośredni. Wskazuje się na jej synonimiczny charakter względem wyrażenia <i>по правде говоря</i> .	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Forum dyskusyjne https://otvet.mail.ru/question/231635800 . [data dostępu: 10.05.2025].	Internauci określają frazę <i>честно говоря</i> jako sygnał częściowego ujawnienia prawdy, często nieprzyjemnej, stosowany w celu wzbudzenia zaufania lub złagodzenia wypowiedzi. Niektórzy wskazują również na jej potencjalnie manipulacyjny charakter.	Częściowe potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Artykuł naukowy Лю Э., <i>Устойчивое выражение «честно говоря» как метатекстовый оператор</i> , [в:] <i>Мир науки, культуры, образования</i> , № 1 (80), Дальневосточный федеральный университет, Владивосток 2020, с. 341-342.	Autor interpretuje wyrażenie jako operator metatekstowy, obecny głównie w kontekstach z negacją, oceną i wyrażeniem postawy nadawcy. Pełni funkcję łagodzenia komunikatu i sygnalizuje autorefleksję mówiącego.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

если честно		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Baza synonimów sinonim.org/t/если честно (data dostępu: 10.05.2025).	Na wskazanej stronie fraza <i>если честно</i> interpretowana jest jako wyrażenie wprowadzające, służące wyrażeniu szczerości lub zapowiedzi, że następne stwierdzenie może mieć charakter	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

	zaskakujący, osobisty lub bezpośredni. Uwzględnione zostały również synonimy <i>откровенно говоря</i> oraz <i>честно говоря</i> .	
Artykuł internetowy https://heroine.ru/7-fraz-kotorye-govoryat-lzheczy-chtoby-zastavit-poverit-v-ih-slova/?ysclid=mdbfbnk8rw690266048 [data dostępu: 10.05.2025].	Fraza <i>если честно</i> została zinterpretowana jako strategia retoryczna, często stosowana przez osoby kłamiące w celu zwiększenia wiarygodności wypowiedzi i ukrycia prawdy.	Brak potwierdzenia obserwacji w źródłach korpusowych.
Artykuł internetowy https://calc-kirpich.ru/stat/polezno/znacenie-frazy-esli-cestno?ysclid=mdbfcrhlp377827177 [data dostępu: 10.05.2025].	Fraza została zinterpretowana jako sygnał wyrażenia szczerości i intymności wypowiedzi, akcentuje odmienną opinię od wcześniejszych sądów lub norm, a także może służyć budowaniu zaufania i emocjonalnej bliskości. W zależności od kontekstu może jednak przybierać wydźwięk ironiczny lub podważać prawdziwość wcześniejszych treści.	Częściowe potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>откровенно говоря</i>		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Gramota.ru gramota.ru/poisk?query=откровенно говоря&mode=slovari&ysclid=mdbfedw060601347209 [data dostępu: 10.05.2025].	Fraza została określona jako środek służący wyrażeniu szczerzej i otwartej opinii, wykorzystywany do formułowania wypowiedzi o charakterze poufnym lub osobistym.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Wpis na stronie https://telegra.ph/CHto-takoe-otkrovenno-govorya-12-24?ysclid=mdbfhlcw2c579880075 [data dostępu: 10.05.2025].	Fraza <i>откровенно говоря</i> jest interpretowana jako sygnał szczerości i otwartości, podkreślający autentyczność wypowiedzi i budujący zaufanie między rozmówcami.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Rosyjski słownik jednojęzyczny <i>Большой академический словарь русского языка. Российская академия наук, Институт</i>	Fraza <i>откровенно говоря</i> została przedstawiona jako formuła metatekstowa służąca zapowiedzi subiektywnej, szczerzej opinii.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

лингвистических исследований. Москва–Санкт-Петербург: Наука, 2013.		

<i>no правде говоря</i>		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Gramota.ru https://gramota.ru/biblioteka/spravochniki/spravochnik-po-punktuiatsii/po-pravde-govorya-skazat? [data dostępu: 10.05.2025].	Fraza <i>no правде говоря</i> podkreśla szczerość i prawdziwość wypowiedzi. Może występować także w wariancie <i>no правде сказать</i> .	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
Słownik frazeologiczny phraseologiya.academic.ru/805/пo_правде_говоря? [data dostępu: 10.05.2025].	Fraza <i>no правде говоря</i> została określona jako wyrażenie potoczne, używane w celu podkreślenia szczerości i prawdziwości wypowiedzi. Jest synonimiczna m.in. <i>честно говоря</i> .	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

Materiały pozakorpusowe, pochodzące głównie ze słowników oraz artykułów internetowych, nie tylko potwierdzają wcześniejsze obserwacje dokonane na podstawie danych korpusowych, lecz także dostarczają dodatkowych, istotnych informacji dotyczących użycia i interpretacji analizowanych jednostek. Podobnie jak w przypadku polskiego wyrażenia *szczerze mówiąc*, opracowania rosyjskie koncentrują się przede wszystkim na ogólnej charakterystyce tych form, wskazując na ich funkcję wyrażania opinii w sposób szczerzy i otwarty. W niektórych źródłach uwzględniono również ich rolę w łagodzeniu krytyki lub przekazywaniu treści potencjalnie zaskakujących. Na szczególną uwagę zasługuje fakt, że część opracowań internetowych interpretuje niektóre z tych jednostek jako elementy wypowiedzi o charakterze manipulacyjnym – wskazuje się na ich możliwe użycie przez nadawcę w sytuacjach nieszczerości, w celu wywarcia wpływu lub budowania fałszywego wrażenia autentyczności. Taka interpretacja może podawać w wątpliwość odbiór tych jednostek jako szczerych i autentycznych, a także nadawać im funkcję ironii bądź kwestionowania wcześniejszych treści wypowiedzi. Zgodnie z wcześniejszymi ustaleniami, część źródeł klasyfikuje wybrane jednostki jako wzajemnie synonimiczne, co potwierdza ich funkcjonalne podobieństwo. Warto również podkreślić, że w jednym z opracowań jednostka *честно говоря* — uznana wcześniej za najbardziej trafny rosyjski odpowiednik — została scharakteryzowana jako operator metatekstowy, co dodatkowo potwierdza jej strukturalne i

funkcjonalne podobieństwo do polskiego *szczerze mówiąc*. Analiza danych pozakorpusowych pozwala zatem na potwierdzenie wcześniejszych ustaleń opartych na materiale korpusowym, wskazując na wysoką zgodność funkcjonalną rosyjskich odpowiedników z polskim wyrażeniem *szczerze mówiąc*. Uwzględniając powyższe obserwacje, możliwe jest zaproponowanie następującej gradacji potencjalnych par przekładowych, opartej na stopniu ich adekwatności względem translandu oraz częstości występowania w korpusach:

1. *szczerze mówiąc* → *честно говоря*,
2. *szczerze mówiąc* → *если честно*,
3. *szczerze mówiąc* → *откровенно говоря*.
4. *szczerze mówiąc* → *по правде говоря*.

Końcowy etap analizy polega na zastosowaniu metody przekładu odwrotnego, opartego na danych pochodzących z rosyjsko-polskiej części korpusu InterCorp v16ud oraz z równoległego korpusu NKJR:

TRANSLAND ROSYJSKI	TRANSLAT POLSKI	FREKWENCJA
<i>честно говоря</i>	<i>szczerze mówiąc</i>	614 poświadczeń (15%)
	<i>prawdę mówiąc</i>	147 poświadczeń (3,7%)
	<i>szczerze powiedziałwszy</i>	30 poświadczeń (0,76%)
<i>если честно</i>	<i>szczerze mówiąc</i>	441 poświadczeń (14,6%)
	<i>prawdę mówiąc</i>	108 poświadczeń (3,5%)
	<i>jeśli mam być szczerzy</i>	98 poświadczeń (3,2%)
<i>откровенно говоря</i>	<i>szczerze mówiąc</i>	89 poświadczeń (20,3%)
	<i>prawdę mówiąc</i>	22 poświadczenia (5%)
	<i>szczerze powiedziałwszy</i>	8 poświadczeń (1,8%)
<i>по правде говоря</i>	<i>prawdę mówiąc</i>	159 poświadczeń (14%)
	<i>szczerze mówiąc</i>	73 poświadczenia (6,4%)
	<i>po prawdzie</i>	22 poświadczenia (1,9%)

Analiza przekładu odwrotnego potwierdziła, że spośród badanych rosyjskich jednostek fatycznych najczęściej wybieranym i zarazem najbardziej odpowiednim ekwiwalentem w

języku polskim pozostaje konstrukcja *szczerze mówiąc*. Jednostka ta pojawia się systematycznie jako podstawowy translat zarówno dla *честно говоря; если честно*, jak i *откровенно говоря*, co świadczy o jej wysokim stopniu ugruntowania w polskim dyskursie i dużej polifunkcyjności. Na drugim miejscu pod względem frekwencji występuje *prawdę mówiąc*, który w przypadku wyrażenia *no правде говоря* stanowi dominujący odpowiednik, najprawdopodobniej ze względu na jego zgodność strukturalną i semantyczną. Należy podkreślić, iż w przypadku jednostki *честно говоря* zarejestrowano także wystąpienia przekładu *szczerze powiedziawszy*, co potwierdza wcześniejsze ustalenia dotyczące wariantywności tej formy i jej bliskiego powiązania z główną strukturą *szczerze mówiąc*. Podobnie potwierdziły się obserwacje dotyczące tłumaczenia *если честно*, które często przekładane jest jako *jeśli tam być szczerzy*, co można tłumaczyć formalną spójnością wynikającą z identycznego komponentu inicjalnego (*если / jeśli*). Warto odnotować, że mimo obecności w polszczyźnie form takich jak *otwarcie mówiąc* czy *mówiąc otwarcie*, nie są one praktycznie stosowane jako translaty dla wyrażenia *откровенно говоря*. Interesującym zjawiskiem jest również obecność skróconej, potocznej formy *po prawdzie* jako tłumaczenia dla *no правде говоря*. Pomimo faktu, iż liczba jej poświadczeń jest niewielka, jej występowanie może świadczyć o procesie eliptyzacji fraz metatekstowych w języku potocznym. Na podstawie uzyskanych danych można stwierdzić, że wszystkie analizowane jednostki rosyjskie realizują zbliżone funkcje pragmatyczne do form wyjściowych w języku polskim, jednak szczególnie wysoką zgodność – zarówno pod względem formalnym, jak i pragmatycznym – wykazują jednostki *честно говоря* oraz *если честно*.

W kolejnym etapie analizy podjęta zostanie analogiczna próba identyfikacji czeskiego odpowiednika dla polskiej jednostki *szczerze mówiąc*. W tym celu ponownie wykorzystano dane z polsko-czeskiej części korpusu InterCorp v16ud, stosując jednakowe założenia metodologiczne jak w poprzednich etapach badania. Kryteria wyszukiwania zostały dostosowane tak, aby jak najwierniej odwzorować formę oraz funkcję jednostki wyjściowej. Na tej podstawie wyodrębniono poniższy zbiór przykładów tłumaczeniowych, w których przeanalizowano możliwe czeskie translanty wraz z informacją o częstotliwości ich występowania:

INTERCORP

☐  adams-stoparuv_pruvodc	Wydaje mi się, szczerze mówiąc , że problem polega na tym, iż wy nigdy nie znaliście naprawdę pytania.	adams-stoparuv_pruvodc	Ale abych był uprząmny, myslím si, že problém spočívá v tom, že jste nikdy pořádně nevěděli, jak zní otázka.
☐  _EUROPART	Szczerze mówiąc , coraz bardziej zastanawiam się, co - oprócz prowadzenia niekończących się rozmów - robi się w Parlamencie Europejskim i w UE, a także co, oprócz gadania, udało nam się w praktyce uczynić, żeby poprawić sytuację Romów w Europie.	_EUROPART	Uprímně řečeno se sama sebe stále častěji ptám, zda kromě mluvení se v Evropském parlamentu i v rámci celé Evropské unie děje ještě něco jiného, a co jsme, kromě mluvení, konkrétně dokázali udělat pro zlepšení situace Romů v Evropě.
☐  _SUBTTILES	Szczerze mówiąc wolalbym ogarnąć jakiegoś placaka.	_SUBTTILES	Abych řekl pravdu, mám spíš chuť na hovně.
☐  ishiguro-neopoustej_me	Szczerze mówiąc , nie wiem nawet, czy to ona pierwsza wspomniała o zwierzętach.	ishiguro-neopoustej_me	Po pravdě řečeno nemůžu ani s určitostí říct, že se o těch zvířatech jako první zmínila ona.

Ekwiwalent	Liczba poświadczeń danego ekwiwalentu	Procentowy udział w ogólnej liczbie poświadczeń
<i>abych byl upřímný/abych byla upřímná</i>	513 poświadczeń	9,3%
<i>upřímně řečeno</i>	498 poświadczeń	9%
<i>abych řekl pravdu/abych řekla pravdu/ abych pravdu řekl/a</i>	208 poświadczeń	2,4%
<i>po pravdě řečeno/popravdě řečeno</i>	77 poświadczeń	1,3%

Przeprowadzona próba wyszukania odpowiedników translatów czeskich wykazała, iż najczęściej używanym odpowiednikiem jednostki *szczerze mówiąc* w zasobach korpusowych jest forma *abych byl upřímný* (lub w wariancie żeńskim *abych byla upřímná*), dla których odnotowano łącznie 513 poświadczeń. Można jednak wstępnie przypuszczać, że konstrukcja ta wydaje się znaczeniowo bliższa polskiemu wyrażeniu *jeśli mam być szczerzy*. Nie bez znaczenia pozostaje również fakt, iż nieznacznie mniej poświadczeń – 498 – uzyskano dla przekładu dosłownego *upřímně řečeno*, który występował wyłącznie w tej formie, bez wariantywności składniowej czy leksykalnej. Kolejnym odpowiednikiem, odnotowanym łącznie w 208 poświadzeniach, jest translat *abych řekl pravdu* wraz z jego wariantami: żeńską formą *abych řekla pravdu* oraz wersjami z przestawionym szykiem składniowym – *abych pravdu řekl* i *abych pravdu řekla*. Wszystkie te realizacje stanowią bardziej rozbudowaną konstrukcję składniową, zachowującą jednak jednolitą funkcję semantyczno-pragmatyczną. Ostatnimi z wyekscerpowanych jednostek tłumaczeniowych stanowią formy *po pravdě řečeno* oraz *popravdě řečeno*, z których pierwsza wykazała wyższą frekwencję (45 poświadczeń), podczas gdy druga wystąpiła w 22 przypadkach, co łącznie daje 77 poświadczeń. Wstępna analiza sugeruje, iż jednostki te są znaczeniowo bliższe polskiemu wyrażeniu *prawdę mówiąc*. Obie konstrukcje wykazują większe podobieństwo znaczeniowe do polskiego wyrażenia *prawdę*

mówiąc. Weryfikacja w dodatkowych źródłach potwierdziła, że mimo różnic zapisu, obie formy funkcjonują w identyczny sposób i przekazują ten sam sens, a rozbieżność ma charakter wyłącznie ortograficzny, co uzasadnia ich łączne traktowanie w dalszej analizie. W związku z powyższym, wszystkie zidentyfikowane jednostki zostaną poddane pogłębionej analizie semantycznej w wybranych czeskich korpusach jednojęzycznych: Araneum Bohemicum Maximum, czeskiej części InterCorp v16ud oraz ORAL v1. Celem analizy pozostaje weryfikacja, czy wyekscerpowane jednostki realizują te same funkcje pragmatyczne, co jednostka wyjściowa w języku polskim oraz jej odpowiedniki w języku rosyjskim. Potwierdzenie takiej zgodności umożliwiłoby ustanowienie trójjęzycznej triady przekładowej, możliwej do zastosowania w analogicznych sytuacjach komunikacyjnych. Poniżej przedstawiono zestawienie zidentyfikowanych znaczeń analizowanych jednostek, opatrzone przykładami pochodzącymi z materiału konkordancyjnego:

EKWIWALENT	ZNACZENIE	PRZYKŁAD UŻYCIA
1. <i>abych byl upřímný</i>	Zaznaczenie zgodności wypowiedzi z rzeczywistym przekonaniem mówiącego. Mówiący podkreśla, że wypowiedź odzwierciedla jego osobiste zdanie lub pogląd, który uznaje za prawdziwy, a jej celem jest wyrażenie szczerości i otwartości wobec odbiorcy.	- <u><i>Abych byl upřímný</i></u> , nikdy jsem na práci s roboty neviděl nic užitečného, ale každý z nás má svoje představy, že?
2. <i>abych byla upřímná</i>	Przyznanie się do wahania lub wewnętrznego oporu. Wyrażenie sygnalizuje trudność w wypowiedzeniu danej treści. Mówiący daje do zrozumienia, że waha się bądź powstrzymuje od jej ujawnienia	- <i>To se mnou sex předtím byl tak špatný?</i> - <i>Ne, ne, ne. Byl dobrý. Ale tohle...</i> - <i>Myslím, tohle ses očividně od někoho naučila.</i> - <i>No, <u>abych byla upřímná</u>...</i> - <i>Nemusíš to vysvětlovat. Spala jsi s jiným klukem.</i>
3. <i>abych byl upřímný</i>	Zapowiedź sprzeciwu, opinii krytycznej lub niepopularnej. Jednostka pełni funkcję amortyzującą – przygotowuje odbiorcę na wypowiedź, która może być niezgodna z jego oczekiwaniami,	- <i>Nejlepší autopůjčovna? <u>Abych byl upřímný</u>, nejlepší autopůjčovna ve skutečnosti neexistuje. V posledních pár let půjčování aut tady v</i>

	wzbudzać sprzeciw lub zawierać treść trudną do przyjęcia.	<i>Chorvatsku, měl jsem dobré a špatné zkušenosti [...].</i>
4. <i>abych byl upřímný</i>	Podkrešlení zgodnosti opisu z rzeczywistością w wypowiedzi neutralnej lub sprawozdawczej. Mówiący nie ocenia i nie wyraża emocji, lecz zaznacza, że przedstawiany stan rzeczy jest zgodny z faktami, jak je postrzega.	- <i>Naši spotřebitelé požadují - a také naši zemědělci, a <u>abych byl upřímný</u>, i my -, abychom měli naprosto stejné normy vysledovatelnosti v Brazílii i ve všech třetích zemích.</i>
5. <i>abych byl upřímný</i>	Przyznanie się do niewiedzy lub braku stanowiska. Mówiący sygnalizuje brak wiedzy, niepewność lub niemożność sformułowania jednoznacznej opinii. Może mieć charakter samoobronny lub dystansujący.	- <i>Carter se opřel do křesla, potřásl hlavou, zkrřížil svalnaté paže na prsou a zahleděl se do tmy. - „Nějaký problém?“ zeptala se Amanda. - „<u>Abych byl upřímný</u>, nevím, jak to mám brát.“</i>
6. <i>abych byl upřímný</i>	Przyznanie się do zaskoczenia. Mówiący sygnalizuje, że odbiór danej sytuacji, informacji lub zdarzenia był sprzeczny z przewidywaniami mówiącego. Może pełnić funkcję wyrażającą autentyczne zdziwienie, konsternację lub niedowierzanie – zarówno pozytywne, jak i negatywne.	- <i><u>Abych byl upřímný</u>, jsem překvapený, že to se mnou udělala. (To nedává smysl). Ale člověče, kašli na to.</i>
7. <i>abych byl upřímný</i>	Przyznanie się do winy lub błędu. Jednostka służy do zasygnalizowania autorefleksji i przyjęcia odpowiedzialności za niepowodzenie, pomyłkę lub niewłaściwe postępowanie. Ma charakter samokrytyczny.	- <i>Moc dobře víte, že je zde zakázáno mít benzinový zapalovač. - I když v něm žádný benzin není, pane řediteli? - I tak... pravidla jsou striktní [...]. - <u>Abych byl upřímný</u>, pane řediteli, zapomněl jsem ho v kapse.</i>

EKWIWALENT	ZNACZENIE	PRZYKŁAD UŽYCIA
1. <i>upřímně řečeno</i>	Zaznacení zgodnosti wypowiedzi z rzeczywistym przekonaniem mówiącego . Mówiący podkreśla, že wypowiedź odzwierciedla jego osobiste zdanie lub pogląd, který uznaje za prawdziwy, a jej celem jest wyrażenie szczerości i otwartości wobec odbiorcy.	- <i>Teprve teď, díky těm příšerným věcem, které zavinil, jsem se dověděl, jaký je mezi námi příbuzenský vztah. Byl to také vnuk Panchy Garcíové, sestry Pedra Segunda, na kterou si <u>upřímně řečeno</u> už vůbec nevzpomínám.</i>
2. <i>upřímně řečeno</i>	Priznání si do wahania lub wewnętrznego oporu . Wyrażenie sygnalizuje trudność w wypowiedzeniu danej treści. Mówiący daje do zrozumienia, že waha się bądź powstrzymuje od jej ujawnienia	- <i>Jakousi představu? Máme přesnou představu. Proč tě to zajímá? - <u>Upřímně řečeno</u>, ani sama nevím. Já jen ... nějak mě to přitahuje a to obvykle znamená, že mám pomoci.</i>
3. <i>upřímně řečeno</i>	Zapowiedź sprzeciwu , opinii krytycznej lub niepopularnej . Jednostka pełni funkcję amortyzującą – przygotowuje odbiorcę na wypowiedź, która może być niezgodna z jego oczekiwaniami, wzbudzać sprzeciw lub zawierać treść trudną do przyjęcia.	- <i>Ano. Zpíváš si ním a zastrčíš ho do krabice. Je to netopýr! - <u>Upřímně řečeno</u>, myslím, že je to trochu bláznivé. Změňme téma.</i>
4. <i>upřímně řečeno</i>	Podkreślenie zgodności opisu z rzeczywistością w wypowiedzi neutralnej lub sprawozdawczej . Mówiący nie ocenia i nie wyraża emocji, lecz zaznacza, że przedstawiany stan rzeczy jest zgodny z faktami, jak je postrzega.	- <i>Měl rád matematiku, jenže v té svoje spolužáky dávno o dva ročníky předběhl. O jiné předměty se <u>upřímně řečeno</u> moc nestaral. S prací to bylo něco docela jiného.</i>
5. <i>upřímně řečeno</i>	Priznání si do niewiedzy lub braku stanowiska . Mówiący sygnalizuje brak wiedzy, niepewność lub niemożność sformułowania jednoznacznej opinii. Może mieć charakter samoobronny lub dystansujący.	- <i>Ta lavina ticha je nesnesitelná. Proč jsem si vybral zrovna dnešní den? <u>Upřímně řečeno</u> nevím. Někdy to být musí.</i>
6. <i>upřímně řečeno</i>	Priznání si do zaskoczenia . Mówiący sygnalizuje, že odbiór danej sytuacji, informacji lub zdarzenia był	- <i>No, <u>upřímně řečeno</u>, takový večer jsem nečekala. Ale bylo to</i>

	sprzeczny z przewidywaniami mówiącego. Może pełnić funkcję wyrażającą autentyczne zdziwienie, konsternację lub niedowierzanie – zarówno pozytywne, jak i negatywne.	<i>úžasné. Celé to bylo tak nápadité.</i>
7. <i>upřímně řečeno</i>	Przyznanie się do winy lub błędu. Jednostka służy do zasygnalizowania autorefleksji i przyjęcia odpowiedzialności za niepowodzenie, pomyłkę lub niewłaściwe postępowanie. Ma charakter samokrytyczny.	- <u>Upřímně řečeno</u> , <i>cítala jsem se provinile. Na prvním rande jsem měla ty vyzívavé šaty, příliš brzo jsem se s ním vyspala.</i>

EKWIWALENT	ZNACZENIE	PRZYKŁAD UŻYCIA
1. <i>abych řekl pravdu</i>	Zaznaczenie zgodności wypowiedzi z rzeczywistym przekonaniem mówiącego. Mówiący podkreśla, że wypowiedź odzwierciedla jego osobiste zdanie lub pogląd, który uznaje za prawdziwy, a jej celem jest wyrażenie szczerości i otwartości wobec odbiorcy.	- <u>Abych řekl pravdu</u> je mi jedno jaké má holka vlasy, ale záleží mi na tom jaká je uvnitř. Samozřejmě mě musí přitahovat a být pro mě atraktivní.
2. <i>abych řekl pravdu</i>	Przyznanie się do wahania lub wewnętrznego oporu. Wyrażenie sygnalizuje trudność w wypowiedzeniu danej treści. Mówiący daje do zrozumienia, że waha się bądź powstrzymuje od jej ujawnienia	- <u>Abych řekl pravdu...</u> Víte, ona... Když mi bylo deset, ona... Ona a můj otec se rozhodli, že jim to spolu nevyjde. A někoho potkala a myslím,... že tím to skončilo.
3. <i>abych řekl pravdu</i>	Zapowiedź sprzeciwu, opinii krytycznej lub niepopularnej. Jednostka pełni funkcję amortyzującą – przygotowuje odbiorcę na wypowiedź, która może być niezgodna z jego oczekiwaniami, wzbudzać sprzeciw lub zawierać treść trudną do przyjęcia.	- Dnes bylo nejdůležitější vyhodnocení dat a srovnání chování nového motoru. <u>Abych řekl pravdu</u> , čekal jsem mnohem víc, protože na top speedu je rychlejší jen asi o čtyři kilometry za hodinu.

4. <i>abych řekl pravdu</i>	Podkrešení zgodności opisu z rzeczywistością w wypowiedzi neutralnej lub sprawozdawczej. Mówiący nie ocenia i nie wyraża emocji, lecz zaznacza, że przedstawiany stan rzeczy jest zgodny z faktami, jak je postrzega.	<i>Přísný pohled upíral na Trémaona velice nerad. [...] "No naštěstí," pokračoval, jsem našel nějaké údaje jinde. <u>Abych řekl pravdu,</u> přímo ve své skromné knihovničce.</i>
5. <i>abych řekl pravdu</i>	Przyznanie się do niewiedzy lub braku stanowiska. Mówiący sygnalizuje brak wiedzy, niepewność lub niemożność sformułowania jednoznacznej opinii. Może mieć charakter samoobronny lub dystansujący.	<i>- Díky za reakce, nicméně jsem se stále nedozvěděl, proč tedy krmit mraženým? Jde o to, aby mé spíše masožravé ryby měly dostatek proteinu, či balastních látek, ale <u>abych řekl pravdu,</u> tak nevím co to jsou balastní látky.</i>
6. <i>abych řekl pravdu</i>	Przyznanie się do zaskoczenia. Mówiący sygnalizuje, że odbiór danej sytuacji, informacji lub zdarzenia był sprzeczny z przewidywaniami mówiącego. Może pełnić funkcję wyrażającą autentyczne zdziwienie, konsternację lub niedowierzenie – zarówno pozytywne, jak i negatywne.	<i>- <u>Abych řekl pravdu,</u> tak je pro mě neuvěřitelné, že počtvrté v řadě zde stojím. Je to něco, čemu nemohu uvěřit.</i>
7. <i>abych řekl pravdu</i>	Przyznanie się do winy lub błędu. Jednostka służy do zasygnalizowania autorefleksji i przyjęcia odpowiedzialności za niepowodzenie, pomyłkę lub niewłaściwe postępowanie. Ma charakter samokrytyczny.	<i>- A o tobě. Camille... Chápu, proč mě nenávidíš. <u>Abych řekl pravdu,</u> udělal jsem nějaké hrozné věci. Ale lži tvého strýce byly řečeny, aby tě ochránil.</i>

EKWIWALENT	ZNACZENIE	PRZYKŁAD UŻYCIA
1. <i>Po pravdě řečeno</i>	Zaznaczenie zgodności wypowiedzi z rzeczywistym przekonaniem mówiącego. Mówiący podkreśla, że wypowiedź odzwierciedla jego osobiste zdanie lub pogląd, który uznaje za	<i>[...] <u>Po pravdě řečeno,</u> moc nevěřím – když chodíš k psychiatrovi druhý rok, ještě si na něco takového nevzpomínáš, – ale vy neznáte závislost na</i>

	prawdziwy, a jej celem jest wyrażenie szczerości i otwartości wobec odbiorcy. .	<i>neochvějném, metafyzickém zlu, kdy na vás nic nezáleží.</i>
2. <i>Po pravdě řečeno</i>	Przyznanie się do wahania lub wewnętrznego oporu. Wyrażenie sygnalizuje trudność w wypowiedzeniu danej treści. Mówiący daje do zrozumienia, że waha się bądź powstrzymuje od jej ujawnienia	- <i>Neměla bych být taková. Co mám dělat? Poučte mě, otče. - Omlouvám se, Mi Young. Ale nemůžu. - Co? - Po pravdě řečeno... nejsem kněz. Protože jste si myslela, že jsem kněz, pokaždé, když jsem vás potkal, nechtěl jsem vás zklamat.</i>
3. <i>Po pravdě řečeno</i>	Zapowiedź sprzeciwu, opinii krytycznej lub niepopularnej. Jednostka pełni funkcję amortyzującą – przygotowuje odbiorcę na wypowiedź, która może być niezgodna z jego oczekiwaniami, wzbudzać sprzeciw lub zawierać treść trudną do przyjęcia.	- <i>Chceme -li v ambulantní péči hovořit o slušné úrovni, musíme pro to něco skutečně aktivně dělat.. A o to se myslím viditelně snažíme téměř 20 let. Po pravdě řečeno, při podání žádosti jsem čekal vcelku bezproblémovou podporu, než hledání důvodů proč to nejde.</i>
4. <i>Po pravdě řečeno</i>	Podkreślenie zgodności opisu z rzeczywistością w wypowiedzi neutralnej lub sprawozdawczej. Mówiący nie ocenia i nie wyraża emocji, lecz zaznacza, że przedstawiany stan rzeczy jest zgodny z faktami, jak je postrzega.	- <i>Slečna Melisandra chodila s maminkou do školy, jak do střední, tak do vyšší v Seminary Hill. Po pravdě řečeno byla o čtyři roky níž než maminka.</i>
5. <i>Po pravdě řečeno</i>	Przyznanie się do niewiedzy lub braku stanowiska. Mówiący sygnalizuje brak wiedzy, niepewność lub niemożność sformułowania jednoznacznej opinii. Może mieć charakter samoobronny lub dystansujący.	- <i>Když za týden nic neudělám, nedá mi nic a ještě mi vyhubuje. Přitom nepracuji pro peníze. - Pro co tedy? Nevím. Po pravdě řečeno nevím skoro nic.</i>
6. <i>Po pravdě řečeno</i>	Przyznanie się do zaskoczenia. Mówiący sygnalizuje, że odbiór danej sytuacji, informacji lub zdarzenia był sprzeczny z przewidywaniami mówiącego. Może pełnić funkcję	- <i>Byla to taková radost z vítězství nebo úleva, že to máš za sebou? LH: Určitě obojí, ale radost z vítězství byla tentokrát</i>

	wyrażającą autentyczne zdziwienie, konsternację lub niedowierzanie – zarówno pozytywne, jak i negatywne.	<i>větší. Po pravdě řečeno jsem ho moc nečekal...</i>
7. <i>Po pravdě řečeno</i>	Przyznanie się do winy lub błędu. Jednostka służy do zasygnalizowania autorefleksji i przyjęcia odpowiedzialności za niepowodzenie, pomyłkę lub niewłaściwe postępowanie. Ma charakter samokrytyczny.	- <i>Potřebovala jsem vše vědět, všude být, vše mít pod kontrolou, být lepší než ostatní. Nic z jmenovaného však nebylo vědomé. Po pravdě řečeno, musela jsem být nesnesitelná.</i>

EKWIWALENT	FREKWENCJA
<i>abych řekl pravdu</i>	12 193 poświadczenia
<i>upřímně řečeno</i>	11 262 poświadczenia
<i>po pravdě řečeno</i>	4 665 poświadczeń
<i>abych byl upřímný</i>	4 227 poświadczeń

Przeprowadzona szczegółowa analiza materiału korpusowego pozwoliła stwierdzić, że wszystkie wyodrębnione czeskie jednostki potencjalnie odpowiadające polskiemu wyrażeniu *szczerze mówiąc* realizują zbliżone funkcje pragmatyczno-komunikacyjne. Badane frazy występują najczęściej w pozycji inicjalnej lub pełnią rolę środków podtrzymujących kontakt i spójność wypowiedzi. Warto jednak zwrócić uwagę na pewną nieoczywistą tendencję frekwencyjną: choć forma *abych byl upřímný* najczęściej pojawiała się jako odpowiednik tłumaczeniowy jednostki wyjściowej, to w zasobach jednojęzycznych posiada najniższą frekwencję spośród analizowanych wyrażen (4 227 poświadczeń). Zdecydowanie najwyższą frekwencję uzyskała jednostka *abych řekl pravdu* (12 193 wystąpienia), a niewiele niższą – *upřímně řečeno* (11 262 poświadczenia). Trzecie miejsce pod względem częstości zajmuje *po pravdě řečeno*, z liczbą 4 665 użyc. Na obecnym etapie analizy można z dużym prawdopodobieństwem przyjąć, że wszystkie formy stanowią trafne odpowiedniki tłumaczeniowe, przy czym ze względu zarówno na częstość użycia, jak i stopień funkcjonalnej zgodności z jednostką polską, szczególną uwagę należy poświęcić formom *abych řekl pravdu* oraz *upřímně řečeno*.

Ostatnim etapem przed ostatecznym ustaleniem par przekładowych jest analiza danych pozakorpusowych. W kolejnej części pracy zaprezentowano wyniki przeszukiwań,

koncentrując się na sposobach interpretacji wskazanych jednostek oraz przypisywanych im funkcjach w konkretnych kontekstach komunikacyjnych:

<i>abych byl upřímný</i>		
ŽRÓDŁO	INFORMACJA	WNIOSEK
Sujin M.Kim, <i>Lingvistické hedge v češtině a angličtině</i> , Bakalařská práce, Univerzita Karlova, Praha 2018, s. 9	W niniejszym opracowaniu fraza <i>abych byl upřímný</i> pojawia się jako przykład środka (ang. hedge) służącego do złagodzenia siły wypowiedzi.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>upřímně řečeno</i>		
ŽRÓDŁO	INFORMACJA	WNIOSEK
Słownik Lingea https://slovníki.lingea.pl/czesko-polski/upřímně [data dostępu: 10.05.2025].	Fraza przetłumaczona jako „szczerze mówiąc”	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>abych řekl pravdu</i>		
ŽRÓDŁO	INFORMACJA	WNIOSEK
Nový encyklopedický slovník češtiny https://www.czechency.org/slovník/HEDGE? [data dostępu: 10.05.2025].	Fraza należy do tzw. <i>pojistkových vět</i> – czyli do „zwrotów zabezpieczających” stosowanych, gdy wypowiedź dotyczy tematów społecznie tabuizowanych lub osobliwie niewygodnych dla adresata. Pełni funkcję osłabiania ilokucyjnej siły wypowiedzi.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
F. Čermák, <i>Slovník české frazeologie a idiomatiky 4. Výrazy větné</i> , Nakladatelství LEDA, Praha 2009, s. 713.	Fraza została interpretowana jako środek służący do uzupełnienia wypowiedzi o subiektywny sąd mówiącego, często dotyczący nieprzyjemnej lub dotąd przemilczanej kwestii, z naciskiem na jej zgodność z rzeczywistością.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

<i>po pravdě řečeno</i>		
ŹRÓDŁO	INFORMACJA	WNIOSEK
Glosbe https://pl.glosbe.com/cs/pl/po-pravde-reno [data dostępu: 10.05.2025].	Fraza przetłumaczona jako „prawdę mówiąc”.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.
F. Čermák, <i>Slovník české frazeologie a idiomatiky 4. Výrazy větné</i> , Nakladatelství LEDA, Praha 2009, s. 713.	Fraza została interpretowana jako środek służący do uzupełnienia wypowiedzi o subiektywny sąd mówiącego, często dotyczący nieprzyjemnej lub dotąd przemilczanej kwestii, z naciskiem na jej zgodność z rzeczywistością.	Potwierdzenie poprzednich obserwacji w źródłach korpusowych.

W materiałach pozakorpusowych nie odnotowano wielu szczegółowych informacji dotyczących ustalonych czeskich translatów, jednak analiza dostępnych źródeł – głównie słowników tradycyjnych i translatorów internetowych – pozwala potwierdzić wcześniejsze wnioski wyprowadzone na podstawie danych korpusowych. Warto zwrócić uwagę, że większość wyodrębnionych jednostek czeskich można zaklasyfikować jako tzw. *pojistkové věty*, czyli środki służące osłabieniu siły illokucyjnej wypowiedzi, co znajduje potwierdzenie również w jednej z prac licencjackich poświęconych zjawisku *hedge*. W opracowaniu tym liczne jednostki odpowiadające polskiemu wyrażeniu *szczerze mówiąc* uznane zostały za środki dystansujące lub łagodzące komunikaty. Na funkcjonalną bliskość badanych czeskich fraz wskazuje również słownik Čermáka, który traktuje wyrażenia *abych řekl pravdu* i *po pravdě řečeno* jako synonimiczne. Tym samym analiza materiału pozakorpusowego stanowi istotne uzupełnienie wcześniejszych ustaleń, potwierdzając wysoki stopień zgodności funkcjonalnej czeskich odpowiedników z jednostką polską. Uwzględniając wszystkie dotychczasowe obserwacje, możliwe jest zaproponowanie gradacji potencjalnych par przekładowych na podstawie ich stopnia odpowiedniości względem translandu oraz częstości występowania w korpusach:

1. *szczerze mówiąc* → *abych řekl pravdu*,
2. *szczerze mówiąc* → *upřímně řečeno*,
3. *szczerze mówiąc* → *abych byl upřímný*,
4. *szczerze mówiąc* → *po pravdě řečeno*.

Ostatecznym etapem analizy jest zastosowanie metody przekładu odwrotnego, opartej na danych z czesko-polskiej części korpusu InterCorp v16ud.

TRANSLAND CZESKI	TRANSLAT POLSKI	FREKWENCJA
<i>abych řekl pravdu</i>	<i>prawdę mówiąc</i>	523 poświadczenia (30,4%)
	<i>szczerze mówiąc</i>	140 poświadczeń (13,8%)
	<i>prawdę powiedziawszy</i>	18 poświadczeń (1,7%)
<i>upřímně řečeno</i>	<i>szczerze mówiąc</i>	235 poświadczeń (23,3%)
	<i>prawdę mówiąc</i>	88 poświadczeń (5,1%)
<i>abych byl upřímný</i>	<i>szczerze mówiąc</i>	324 poświadczenia (24,9%)
	<i>prawdę mówiąc</i>	42 poświadczenia (3,2%)
	<i>jeśli mam być szczerzy</i>	84 poświadczenia (6,4%)
<i>po pravdě řečeno</i>	<i>prawdę mówiąc</i>	151 poświadczeń (22,8%)
	<i>szczerze mówiąc</i>	51 poświadczeń (7,7%)
	<i>prawdę powiedziawszy</i>	31 poświadczeń (4,6%)

Analiza przekładu odwrotnego z języka czeskiego potwierdziła początkowe obserwacje dotyczące adekwatności poszczególnych translatów, wskazując jako najbardziej trafne odpowiedniki jednostki: *abych byl upřímný* oraz *upřímně řečeno*. Jednostka *abych řekl pravdu*, mimo iż jest najczęściej poświadczana w jednojęzycznych korpusach czeskich, w przekładzie na język polski bywa przeważnie oddawana wyrażeniem *prawdę mówiąc* (523 poświadczenia), co można tłumaczyć obecnością wspólnego komponentu leksykalnego *pravda/prawda*. Wyrażenie *szczerze mówiąc*, które zachowuje bliższą odpowiedniość funkcjonalno-pragmatyczną, pojawia się rzadziej – w 140 przypadkach.

Wśród danych korpusowych odnotowano również jednostkowe przypadki ekwiwalencji z wyrażeniem *prawdę powiedziawszy*, które należy uznać za synonimiczne względem *szczerze mówiąc* oraz wariantywne do re produktu *prawdę mówiąc*. Założenia dotyczące ekwiwalencji jednostki *abych byl upřímný* znalazły potwierdzenie w danych przekładu odwrotnego – wyrażenie to tłumaczone bywa jako *jeśli mam być szczerzy*. Z kolei *po pravdě řečeno* występuje w polskich przekładach przede wszystkim jako *prawdę mówiąc*, rzadziej jako *szczerze mówiąc*.

Uwzględniając cały przebieg analizy – od etapu identyfikacji potencjalnych translatów, poprzez ich weryfikację znaczeniową i frekwencyjną w czeskich korpusach jednojęzycznych, aż po analizę przekładu odwrotnego – możliwe jest wskazanie najbardziej odpowiedniego i funkcjonalnie zgodnego ekwiwalentu w języku czeskim. Najwyższym stopniem zgodności odznacza się jednostka *upřímně řečeno*, następnie *abych řekl pravdu*, zbliżona *abych byl upřímný*, a na końcu *po pravdě řečeno*. Należy jednak podkreślić, iż wszystkie analizowane jednostki realizują zbliżone funkcje komunikacyjne i wykazują wysoki stopień synonimiczności, co zostało wykazane zarówno na podstawie danych ilościowych, jak i analizy jakościowej.

Wyniki przeprowadzonej analizy, opartej na danych korpusowych oraz uzupełnionej weryfikacją w źródłach pozakorpusowych, pozwalają na wskazanie najbardziej trafnych ekwiwalentów tłumaczeniowych dla polskiej jednostki *szczerze mówiąc*. W języku rosyjskim są to konstrukcje *честно говоря* oraz *если честно*, natomiast w języku czeskim odpowiednikami o najwyższej zgodności okazują się *upřímně řečeno* oraz *abych řekl pravdu*. Jednostki te odznaczają się wysokim stopniem zgodności zarówno pod względem struktury formalnej, jak i pełnionych funkcji pragmatycznych, co znajduje potwierdzenie w danych pochodzących z korpusów równoległych oraz analiz jednojęzycznych. Zgromadzony materiał językowy umożliwił również wyodrębnienie głównych funkcji analizowanego reproduktu i jego odpowiedników. Frazy te najczęściej służą wyrażaniu subiektywnej opinii lub przekonania, stanowią deklaracje podkreślające szczerzość bądź autentyczność wypowiedzi, a także pełnią funkcję łagodzenia krytyki lub przyznania się do niewiedzy. Znacznie rzadziej jednostki te pojawiają się w funkcji sygnalizowania wahania (najczęściej zaznaczanego graficznie za pomocą wielokropka) czy też jako środki autokrytyczne, sugerujące przyznanie się do popełnionego błędu. Tendencje te są spójne dla wszystkich trzech języków i zostały potwierdzone w analizie licznych poświadczeń konkordancyjnych. Podsumowując, analizowana jednostka *szczerze mówiąc*, pomimo pozornej znaczeniowej i strukturalnej przezroczystości, jest złożonym, polifunkcyjnym reproduktom fatyczno-pragmatycznym. Może ona inicjować wypowiedź, wzmacniać jej wiarygodność lub służyć podtrzymaniu kontaktu.

POLSKI	ROSYJSKI	CZESKI
<i>1. szczerze mówiąc</i>	<i>1. честно говоря</i>	<i>1. upřímně řečeno</i>
	<i>2. если честно</i>	<i>2. abych řekl pravdu</i>

Przeprowadzona analiza pozwala sformułować szereg wniosków dotyczących zastosowania metody korpusowej w procesie ustalania odpowiedników przekładowych dla jednostek fatycznych i emotywnych. Owa metoda, oparta na wykorzystaniu korpusów równoległych i jednojęzycznych, okazuje się użytecznym narzędziem we wskazywaniu potencjalnych translatów, umożliwiając zarówno wyodrębnienie najczęściej stosowanych konstrukcji tłumaczeniowych, jak i pogłębioną analizę funkcji pragmatycznych w pozyskanych konkordancjach. Zaproponowany ciąg procedur – od identyfikacji jednostki wyjściowej po weryfikację przekładu odwrotnego – może znaleźć zastosowanie również w analizie pozostałych fatyzmów i emotywów, przy zachowaniu analogicznych zasad postępowania badawczego. Analiza frekwencji, uzupełniona badaniem pozakorpusowym, umożliwia gradację odpowiedników i weryfikację ich funkcjonalnej adekwatności względem translandu.

W przypadku badanych fraz – *rób, co chcesz* oraz *szczerze mówiąc* – metoda korpusowa pozwoliła na precyzyjne wskazanie jednostek rosyjskich i czeskich, które wykazują największe podobieństwo semantyczne i pragmatyczne. Należy jednak zauważyć istotne ograniczenie metody korpusowej, polegające na stosunkowo niskiej frekwencji wielu analizowanych jednostek w korpusach równoległych, co wynika z ich przynależności do rejestru języka mówionego, rzadziej reprezentowanego w tego typu zasobach. W konsekwencji nawet przy dobrze opracowanej procedurze, konieczne staje się uzupełnienie analizy o dodatkową weryfikację semantyczno-pragmatyczną oraz weryfikację źródeł pozakorpusowych. Analiza ujawniła ponadto, że funkcje analizowanych wyrażen są często zróżnicowane i niejednoznaczne, co utrudnia jednoznaczne przypisanie ich do konkretnych translatów. Podsumowując, metoda korpusowa skutecznie wspiera proces ustalania par przekładowych, zwłaszcza na etapie wstępnej selekcji odpowiedników. Jej przydatność wzrasta w połączeniu z analizą danych jednojęzycznych oraz weryfikacją w źródłach pozakorpusowych. Przy zachowaniu krytycznego podejścia i świadomości jej ograniczeń, może stanowić solidną podstawę dla dalszych badań translacyjnych, szczególnie w zakresie przekładu jednostek wielofunkcyjnych.

7.2 Metoda audiowizualna

Współczesny przekład audiowizualny (AVT, ang. *audiovisual translation*) stanowi jedną z najdynamiczniej rozwijających się dziedzin translatoryki, uwzględniającą szczególnie typ tekstów – wielosemiotycznych, złożonych z warstwy werbalnej, wizualnej i audytywnej. Specyfika tego rodzaju przekładu wiąże się z koniecznością uwzględnienia nie tylko ograniczeń

technicznych i czasowych, ale również złożonej interakcji między kanałami komunikacyjnymi, co istotnie wpływa na dobór strategii translatorskich (Belczyk 2007: 6). Charakterystyczne dla AVT jest jego głębokie osadzenie w kontekście sytuacyjnym i kulturowym, które w znacznym stopniu determinują formę wypowiedzi oraz dobór ekwiwalentów tłumaczeniowych. We współczesnej typologii wyróżnić możemy trzy podstawowe formy AVT: *napisy filmowe* (subtitling), *lektora* (voice-over) oraz *dubbing*. W przypadku napisów warstwa dźwiękowa oryginału zostaje zachowana, natomiast tekst tłumaczenia jest wyświetlany na ekranie w sposób zredukowany i zsynchronizowany z obrazem. Tego rodzaju przekład wymusza dużą kondensację treści i często prowadzi do utraty elementów ekspresywnych czy kolokwialnych. Lektor natomiast polega na nałożeniu nowej ścieżki dźwiękowej zawierającej tłumaczenie przy jednoczesnym zachowaniu w tle oryginalnej ścieżki fonicznej. Forma ta jednak w sposób istotny ogranicza swobodę tłumacza, a przez monotony, pozbawiony emocji ton lektora utrudnia oddanie naturalnej intonacji i dynamiki dialogu. Z perspektywy niniejszej pracy szczególnie istotny pozostaje dubbing, który – jak podkreśla się w czeskich opracowaniach – należy do najczęściej stosowanych form przekładu filmowego i umożliwia całkowite zastąpienie oryginalnej warstwy fonicznej przekładem w języku docelowym. Dzięki temu możliwe staje się nie tylko pełne dostosowanie treści do realiów językowo-kulturowych, ale także wierne oddanie funkcji fatycznych bądź emotywnych.

Jak zauważają badacze, „cílem dabingu je docílit stejné dynamiky, jakou měl originál – klíčové je dodržet stejné tempo promluvy a vybírat taková překladatelská řešení, která budou v souladu s labiální synchronizací” (Grohová 2024: 13), co wymaga szczególnej dbałości o prozodię, kadencję i rytm języka. Pomimo technicznych ograniczeń, dubbing „má však potenciál velmi zdařile zachytit rejstřík a formu původního sdělení” (Grohová 2024: 13), co czyni go narzędziem szczególnie przydatnym w analizie jednostek o charakterze pragmatycznym. Filmy z dubbingiem pozwalają również na rekonstrukcję warunków komunikacji naturalnej i ułatwiają przyswajanie idiomatycznych konstrukcji językowych. Tym samym ich potencjał może być z powodzeniem wykorzystany również w badaniach nad przekładalnością struktur potocznych i funkcjonalnie nacechowanych.

W kontekście niniejszych rozważań warto również przywołać koncepcję obrazu jako *tertium comparationis* – zewnętrznej kategorii referencyjnej, która może posłużyć jako wspólna płaszczyzna porównawcza między oryginałem a przekładem. Choć w tradycyjnej teorii przekładu audiowizualnego obraz uznawany jest za integralny składnik przekazu, niektórzy badacze wskazują, że w pewnych przypadkach może on pełnić funkcję niezależnego

odniesienia, pozwalającego kontrastować decyzje tłumaczeniowe względem przekazu wizualnego (Bogucicki 2015: 84).

Należy jednak podkreślić, że celem wykorzystania materiałów audiowizualnych w niniejszej pracy nie jest analiza samego procesu tłumaczenia ani ocena jakości przekładu AVT. Przedmiotem zainteresowania nie jest zatem strategia tłumacza, lecz *produkt końcowy* – czyli gotowa wersja dubbingowa, która umożliwia identyfikację par przekładowych na podstawie porównania wersji polskiej, rosyjskiej i czeskiej. Dzięki temu możliwe staje się uchwycenie sytuacyjnych ekwiwalentów przekładowych, odzwierciedlających praktykę przekładu konkretnych jednostek fatycznych i emotywnych, funkcjonujących w kontekście dialogowym i pragmatycznym. Istotną zaletą dubbingu pozostaje jego zdolność do takiej adaptacji tekstu, która pozwala na osiągnięcie wysokiego stopnia naturalności przekazu w języku docelowym. Jak zauważa N. N. Nevskich,

«стратегия дублирования требует адаптации текста на различных уровнях с целью заменить текст оригинала таким образом, чтобы аудиовизуальный продукт создавал впечатление, будто он был исходно создан на языке оригинала» (Невских 2024: 22).

Tego rodzaju integracja werbalnego i wizualnego wymiaru przekazu sprzyja analizie jednostek o funkcji pragmatycznej, pozwalając uchwycić ich realizację nie tylko na poziomie treści, lecz również intonacji i sytuacyjnego osadzenia. Warto jednak zaznaczyć, że omawiana metoda posiada również pewne ograniczenia. Ze względu na sytuacyjny charakter analizy, ustalone pary przekładowe odnoszą się wyłącznie do konkretnego kontekstu, co nie pozwala na uchwycenie pełnej wieloznaczności badanych jednostek. Jak wykazano wcześniej, zarówno fatyzmy, jak i emotywy mogą realizować różne funkcje pragmatyczne, zależnie od warunków użycia, a wyodrębniona na podstawie materiału dubbingowanego para tłumaczeniowa odzwierciedla jedynie jedną z możliwych realizacji danej jednostki. Pomimo faktu, iż dostarcza ona często trafnego przykładu użycia, nie może zostać bezpośrednio przeniesiona do słownika bez dalszej analizy jej pozostałych funkcji. Niemniej jednak, samo uzyskanie odpowiednika poświadczanego w praktyce przekładowej i osadzonego w konkretnym kontekście sytuacyjnym stanowi istotne wsparcie w procesie ustalania ekwiwalentów tłumaczeniowych dla jednostek fatycznych i emotywnych.

Na potrzeby niniejszego badania, w celach egzemplifikacyjnych, zdecydowano się na wykorzystanie dubbingowanego filmu animowanego jako materiału umożliwiającego identyfikację jednostek fatycznych i emotywnych w języku polskim, rosyjskim i czeskim. Filmy animowane, zwłaszcza te adresowane zarówno do dzieci, jak i dorosłych, stanowią niezwykle cenne źródło danych z perspektywy analizy reproduktów języka mówionego.

Charakteryzują się one wysokim stopniem nasycenia żywą, ekspresywną leksyką oraz częstym wykorzystaniem kolokwializmów, zwrotów dialogowych i emocjonalnych. Obecność licznych gier słownych, aluzji kulturowych oraz dynamicznych wymian dialogowych sprzyja analizie zjawisk o charakterze pragmatycznym, takich jak fatyzmy i emotywy. Co istotne, przekład dubbingowy w tego rodzaju produkcjach zazwyczaj charakteryzuje się dużą swobodą językową i kreatywnością, co pozwala uchwycić strategie tłumaczeniowe zachowujące funkcje oryginału. Szczególnie interesującym i reprezentatywnym przykładem jest w tym kontekście film *Shrek* – jedna z najpopularniejszych produkcji animowanych XXI wieku, wyprodukowana przez wytwórnię DreamWorks i wprowadzona na ekrany w 2001 roku. Film ten – skierowany do szerokiego grona odbiorców – zyskał uznanie nie tylko ze względu na walory artystyczne, ale również za sprawą bogatej warstwy językowej i kulturowej. Jak zauważa Karolina Wakulik, *Shrek* operuje wielopiętrowym kodem intertekstualnym, stanowiącym wyzwanie dla tłumacza i wymagającym wysokich kompetencji interkulturowych. Autorka podkreśla, że dla danego filmu charakterystyczny jest wysoki poziom intertekstualności, czyli umiejętności komunikowania się za pomocą odniesień do tekstów kultury, których znaczenie może różnić się w zależności od kontekstu kulturowego (Wakulik 2023: 599–614). Ta złożoność kodów kulturowych i językowych sprawia, że film ten stanowi doskonałe pole badawcze w kontekście analizy przekładu pragmatycznego. Jego zawartość obfituje w naturalne dialogi, wypowiedzi nacechowane emocjonalnie, elementy komiczne i ironiczne, a także liczne przykłady zwrotów potocznych służących podtrzymywaniu kontaktu, co czyni z niego wartościowy materiał w analizie jednostek fatycznych i emotywnych w trzech językach: polskim, rosyjskim i czeskim. W dalszej części badania film *Shrek* posłuży jako materiał analityczny, na podstawie którego wyodrębnione zostaną jednostki fatyczne i ematywne obecne w polskiej wersji dubbingowanej. Następnie przeprowadzona zostanie szczegółowa analiza porównawcza, mająca na celu ustalenie, czy w odpowiadających momentach wersji rosyjskiej i czeskiej zastosowano tłumaczenia sytuacyjne – to znaczy takie, które w naszej ocenie najtrafniej oddają funkcję komunikacyjną oryginalnych wyrażen.

Należy jednak zaznaczyć, że uzyskane trójjęzyczne zestawienia nie mają charakteru ekwiwalencji leksykalnej ani strukturalnej w sensie ścisłym. W większości przypadków mamy do czynienia z ekwiwalentami funkcjonalnymi, które – mimo różnic formalnych, długości, obrazowości czy stopnia konwencjonalizacji – realizują zbliżone funkcje pragmatyczne. Taki stan rzeczy wynika m.in. ze specyfiki przekładu audiowizualnego, który podlega licznym ograniczeniom technicznym, takim jak konieczność dopasowania długości wypowiedzi do obrazu, synchronizacja ruchu warg czy tempo sceny. W konsekwencji tłumacze podejmują

decyzje mające na celu nie tyle zachowanie struktury oryginału, ile jego funkcji komunikacyjnej i stylistycznej w danym kontekście kulturowym. W związku z tym zestawione jednostki należy traktować jako funkcjonalne odpowiedniki sytuacyjne, a nie jako tłumaczenia ekwiwalentne w sensie formalnym. Na potrzeby analizy wybrane zostaną pary przekładowe (polsko–rosyjskie i polsko–czeskie), które następnie poddane zostaną dalszej weryfikacji z wykorzystaniem źródeł korpusowych. Celem tego etapu będzie potwierdzenie adekwatności funkcjonalnej zastosowanych odpowiedników w szerszym kontekście językowym. Należy również podkreślić, że zastosowana metoda pełni funkcję pomocniczą – nie pozwala na uchwycenie całego spektrum znaczeniowego danej jednostki, lecz umożliwia wskazanie jednego z jej możliwych znaczeń sytuacyjnych. Znaczenie to może następnie posłużyć jako punkt odniesienia w dalszych etapach analizy, a także jako materiał uzupełniający przy opracowywaniu słownika dwujęzycznego oraz trójjęzycznego. Poniżej zaprezentowano zestawienie około czterdziestu przykładów jednostek fatycznych i emotywnych wyodrębnionych z pierwszej części filmu *Shrek*, wraz z ich odpowiednikami w wersji rosyjskiej i czeskiej:

L.p.	Czas	Wersja polska	Wersja rosyjska	Wersja czeska
1.	3:50	Gwoli ścisłości	-	Po pravdě
2.	5:26	Zamknij się	Заткнись	-
3.	06:03	Dość, tracimy czas	Ясно, довольно	Tak dost. To myslím stačí
4.	7:22	O, serio?	Неужели?	Neříkej
5.	7:38	Mogę wtrącić?	Можно мне сказать?	Můžu ti něco říct?
6.	9:33	Nie przeszkadza ci to?	-	To ti vůbec nevadí?
7.	9:36	Serio?	Правда?	Vážně?
8.	9:38	Serio, serio.	Правда, правда.	-
9.	10:25	Ty wiesz, że ja też	-	Přesně jako já
10.	10:41	Ależ jasne!	Как хочешь	Ale to víš, že..
11.	10:55	Dobrze, już dobrze	Ну хорошо, хорошо	Dobře, tak jo
12.	13:55	No a kto ma być?	Как ты догадался?	Jak jsi to poznal?
13.	14:26	Na mnie nie patrz	-	Na mě nekoukej
14.	15:34	Tak jest!	Клёво!	Bezva!
15.	16:57	Ugryź się	Вот тебе!	-
16.	17:33	To na co jeszcze czekasz?	Так чего вы ждёте?	Tak na co ještě čekáte?
17.	18:12	Miałem na myśli...	В смысле...	Chtěl jsem říct..
18.	26:21	Ustalmy fakty...	Давай разберёмся...	Tak znova...(?)
19.	26:28	Mam rację?	Я ничего не пропустил?	Říkám to dobře?
20.	26:29	Wiesz, co?	Вот беда	Víš ty co?
21.	29:57	Bo widzisz, jest taka sprawa...	Должен тебе признаться в том, что...	Musím se ti k něčemu přiznat...
22.	30:02	Zaraz, zaraz	-	Tak moment
23.	32:30	Zamknij się	Заткни хлебало	Drž zobák

24.	37:12	Wiesz, że nie o to mi chodzi	Речь вовсе не о том	Ale o to přece nejde
25.	42:39	To wszystko nie tak	Всё неправильно	Všechno je špatně
26.	42:54	Dobre pytanie	Хороший вопрос	Dobrá otázka
27.	44:54	Daj spokój	Эй небойся	Klid, nebojte se
28.	45:38	Daj spokój, żartowałem	Уймись, я пошутил	No tak dělám si srandu
29.	46:56	Wiesz, co ja myślę?	-	Víš, co si myslím?
30.	47:01	Nie, serio?	Нет. Думаешь?	Neříkej, fakt myslíš?
31.	50:42	Bądź co bądź...	В конце концов...	Konec konců...
32.	55:12	Daj spokój	Что, спятил?	Neblázní
33.	58:20	Nie chcę się chwalić...	Неудобно хвастать	Nechci se vyťahovat...
35.	59:59	Co ty opowiadasz?	Не понимаю, о чём ты	Co ty zase plácáš?
36.	01:00:06	Odbiło Ci	Ты взбесился	Zbláznil ses
37.	01:06:17	Muszę Ci coś powiedzieć	Я должна в кое-в чём признаться	Chtěla bych Ti něco říct
38.	01:08:37	I co z tego?	И что с того?	-
39.	01:14:05	Nie bój nic	Небойся	Bez obav
40.	01:15:34	O czym ty znowu bredzisz?	Я тебя непонимаю, о чём ты?	Co to zase plácáš?

Jak można zauważyć, w tabeli przedstawiono konkretne fragmenty dialogów, w których zidentyfikowano funkcjonalnie relewantne jednostki językowe, a następnie wskazano ich realizacje w trzech wersjach językowych. Należy podkreślić, iż w niektórych przypadkach tabela zawiera luki, wynikające z faktu, iż dana jednostka w jednej z wersji językowych nie została właściwie przetłumaczona, a zamiast tego zastąpiona została wypowiedzią odnoszącą się do innego aspektu sytuacji komunikacyjnej (np. pol. *Na mnie nie patrz*, ros. *Я не звал этих муное*; czes. *Na mě nekoukej*). W takich przypadkach można mówić o braku funkcjonalnej zgodności tłumaczenia, co uniemożliwia zakwalifikowanie danej realizacji jako trafnego odpowiednika w danym języku. Niemniej jednak obecność dwóch zbieżnych realizacji w pozostałych językach (np. polskiej i czeskiej) może stanowić wartościowy punkt wyjścia do dalszych poszukiwań odpowiednika w języku trzecim. Z kolei w innych przykładach pojawiają się sytuacje, w których wszystkie trzy wersje zawierają potencjalnie ekwiwalentne jednostki – zarówno pod względem strukturalnym, jak i funkcjonalnym – co pozwala na ich wstępne zakwalifikowanie jako kandydatów na trójjęzyczne układy przekładowe. Należy przy tym uwzględnić, że analizowane wersje językowe charakteryzują się zróżnicowanymi strategiami translatorskimi i w zależności od decyzji tłumacza, niektóre wersje cechuje większy stopień ekspresywności, potoczności bądź przeciwnie – neutralizacji i uogólnienia. Tego rodzaju rozbieżności wynikają zarówno z różnic systemowych między językami, jak i z odmiennych preferencji kulturowych oraz konwencji danego rynku audiowizualnego. W świetle powyższego, prezentowane zestawienie należy traktować jako materiał wyjściowy, który

wymaga dalszej weryfikacji z wykorzystaniem danych korpusowych oraz pozakorpusowych. Dopiero analiza w szerszym kontekście językowym pozwoli zweryfikować adekwatność zaproponowanych odpowiedników oraz zakres pełnionych przez nie funkcji, przy czym zaprezentowane przykłady mogą stanowić gotowy materiał do określenia jednego z możliwych znaczeń analizowanych jednostek.

W celu egemplifikacji zaproponowanej metody analizy, w dalszej części zostaną przedstawione wybrane zestawienia trójjęzyczne, w których możliwe było jednoznaczne wyodrębnienie ekwiwalentów tłumaczeniowych we wszystkich trzech wersjach językowych. Należy podkreślić, że na obecnym etapie nie jest zamiarem przeprowadzenie pełnej analizy wszystkich przykładów zawartych w zestawieniu, lecz jedynie zilustrowanie procesu interpretacji wybranych jednostek na podstawie danych audiowizualnych oraz ich weryfikacji w źródłach korpusowych i pozakorpusowych.

Takim przykładem może być jednostka *Mam rację?*, która w analizowanym materiale pojawia się w wersjach: rosyjskiej *Я ничего не proponyтил?* oraz czeskiej *Říkám to dobře?*. W analizowanym kontekście filmowym jednostka ta, wypowiedziana przez postać osła, pojawia się na końcu wypowiedzi i pełni funkcję pytania, służącego upewnieniu się co do trafności interpretacji zdarzeń. Realizuje ona potrzebę uzyskania akceptacji ze strony rozmówcy, potwierdzenia poprawności własnego rozumienia sytuacji oraz weryfikacji posiadanej wiedzy o przebiegu wydarzeń. Wpisanie tej jednostki wraz ze znakiem zapytania do odpowiednich zasobów korpusowych przyniosło następujące wyniki: w przypadku jednostki polskiej odnotowano 2825 poświadczeń (Araneum Polonicum Maius, polska część InterCorp v16ud oraz NKJP), dla jednostki rosyjskiej – 22 poświadczenia (Araneum Russicum Maius, rosyjska część InterCorp v16ud oraz NKJR), natomiast dla jednostki czeskiej – 53 poświadczenia (Araneum Bohemicum Maximum, czeska część InterCorp v16ud oraz korpus ORAL v1). Wśród zgromadzonych konkordancji wyodrębniono poszczególne przykłady kontekstów, w których jednostka realizuje opisaną powyżej funkcję:

- *No widzisz, ale to nieprawda.. Twój mąż był chory. Był nałogowym hazardzistą. Mam rację?*

- *Nie rozumiem, co się kryje pod terminem homeopatia. Kojarzę ją z ziołolecznictwem, mam rację?*

- *Tak też myślałem. Wspaniale podrabia pan paszporty i sądzę, że jest pan na tyle ostrożny, aby mieć prywatne archiwum swych arcydzieł. Mam rację?*

- Оно нагружает икроножные, бицепсы бедер, ягодичные, квадрицепсы, выпрямители позвоночника. Трапеции, дельтоиды, бицепсы, трицепсы и предплечья, мышцы пресса как стабилизаторы. <u>Я ничего не пропустил?</u>
- Этих запасов хватит если бойцы будут стрелять как у нас принято 2 - 3 раза за службу, из них один 3 патрона перед принятием Присяги. 30 лет и на переплавку? <u>Я ничего не пропустил?</u>
- Но через месяц , после кражи бокса, обвинения были сняты. Кто-то, нашел слабое место, у агента секретной службы. Шантажировал вас и заставил, сделать это. <u>Я ничего не пропустил?</u>
Jak to víte? On přece slíbil, že přijde ještě za života těch, kteří s ním byli, a uvede je do nebeského království. <u>Říkám to dobře?</u>
No... bylo by docela nepříjemné, kdyby se na poslední chvíli objevilo něco v práci a on nemohl jet. - Jo. - Hádám , že by to znamenalo, že by si nemusel prohazovat víkendy a tím by se vám uvolnila sobota. <u>Říkám to dobře?</u>
- Vypadá to, že se to seběhlo přibližně takhle. Vy jste Gómeze poslali do Devota v červnu 1972. <u>Říkám to dobře?</u>

Różnice w liczbie poświadczeń dla analizowanych jednostek są znaczące, co może świadczyć o wysokiej frekwencji i powszechnym użyciu polskiego wyrażenia *Mam rację?* w języku codziennym. Z kolei odpowiedniki rosyjski (*Я ничего не пропустил?*) oraz czeski (*Říkám to dobře?*) pojawiają się rzadziej, co może wynikać zarówno z ich rzadszego użycia w komunikacji potocznej, jak i z ograniczeń samych korpusów, które rzadko w pełni rejestrują jednostki charakterystyczne dla języka mówionego. Mimo tych rozbieżności ilościowych analiza kontekstów użycia potwierdza, że wszystkie wskazane konstrukcje pełnią zbliżoną funkcję – służą wyrażeniu potrzeby potwierdzenia własnych słów lub interpretacji sytuacji przez rozmówcę – i najczęściej występują w pozycji końcowej wypowiedzi. W końcowym etapie analizy zasadne jest odwołanie się do korpusów równoległych – polsko-rosyjskiej i polsko-czeskiej części korpusu InterCorp v16ud oraz korpusu równoległego NKJR – w celu weryfikacji, czy uzyskane z materiału filmowego pary przekładowe znajdują potwierdzenie w rzeczywistej praktyce translacyjnej. Podjęto zatem próbę ustalenia, jakie odpowiedniki tłumaczeniowe jednostki *Mam rację?* występują w korpusach równoległych. Poniżej przedstawiono wyłącznie te translaty rosyjskie i czeskie, które w analizowanych

konkordancjach realizowały tę samą funkcję pragmatyczną, co jednostka zaobserwowana w materiale filmowym:

TRANSLAND POLSKI	TRANSLAT ROSYJSKI	FREKWENCJA
<i>Mam rację?</i>	<i>Я прав?</i>	198 poświadczeń (19,4%)
	<i>Верно?</i>	66 poświadczeń (6,4%)
	<i>He так ли?</i>	51 poświadczeń (5%)

TRANSLAND POLSKI	TRANSLAT CZESKI	FREKWENCJA
<i>Mam rację?</i>	<i>Nemám pravdu?</i>	246 poświadczeń (11,1%)
	<i>Mám pravdu?</i>	151 poświadczeń (6,8%)
	<i>Je to tak?</i>	111 poświadczeń (5%)

Przeprowadzona analiza materiału filmowego w zestawieniu z danymi korpusowymi pozwala na sformułowanie kilku istotnych wniosków. W wersjach dubbingowanych filmu *Shrek* jako odpowiedniki polskiej jednostki *Mam rację?* zastosowano konstrukcje *Я ничего не нпоньстил?* (ros.) oraz *Říkám to dobře?* (czes.), które – mimo różnic strukturalnych – realizują analogiczną funkcję: wyrażają potrzebę potwierdzenia własnej interpretacji i uzyskania aprobaty rozmówcy. Weryfikacja odpowiedników w korpusach równoległych (InterCorp v16ud, NKJR) wykazała jednak, że nie występują one jako typowe translaty polskiej jednostki. W zasobach korpusowych dominują konstrukcje strukturalnie bardziej zbliżone do oryginału, m.in. *Я прав?*; *Верно?*; *He так ли?* w języku rosyjskim oraz *Nemám pravdu?*; *Mám pravdu?*; *Je to tak?* w czeskim. Obserwowana rozbieżność pozwala przypuszczać, iż korpusy częściej rejestrują tzw. pełne (lub formalne) ekwiwalenty, utrzymane w podobnym schemacie gramatycznym i leksykalnym co jednostka wyjściowa. Z kolei przekład filmowy, uwarunkowany rytmem dialogu, synchronizacją obrazu i specyfiką stylu postaci, dopuszcza natomiast większą swobodę i pozwala na użycie rozwiązań sytuacyjnych, mniej typowych, lecz funkcjonalnie trafnych. Zestawienie to potwierdza przyjęte wcześniej założenie, że metoda audiowizualna ujawnia jedynie część możliwych realizacji jednostek i dostarcza przede wszystkim ekwiwalentów pragmatycznych, a nie strukturalnych. Odpowiedniki wyodrębnione z filmu, mimo że nie znajdują potwierdzenia w danych korpusowych, nie tracą na wartości – stanowią przykład funkcjonalnego dopasowania wypowiedzi do konkretnego kontekstu komunikacyjnego. Realizują one przyjęte wcześniej kryteria zgodności sytuacyjnej oraz spójności pragmatycznej, rozumianej jako zachowanie sensu i funkcji komunikacyjnej

jednostki źródłowej w porównywalnych kontekstach języka docelowego. Podsumowując, przeprowadzone badanie pozwoliło zidentyfikować zarówno ekwiwalenty funkcjonalne, jak i pełne bądź częściowe, które mogą stanowić materiał do utworzenia par tłumaczeniowych dla jednostek fatycznych i emotywnych o określonym, sytuacyjnie zdefiniowanym znaczeniu. Jednocześnie wskazane jednostki mogą stanowić punkt wyjścia do dalszych poszukiwań bardziej trafnych odpowiedników, przy czym należy pamiętać, że granice między typami ekwiwalencji pozostają płynne i zależne od kontekstu.

W końcowej części analizy materiału filmowego warto podkreślić, iż dostarcza on również istotnych informacji na temat sposobu przekładu jednego z najbardziej wieloznacznych i zarazem problematycznych pod względem tłumaczeniowym emotywów, jakim jest wyrażenie *daj spokój*. Jak zasygnalizowano już we wcześniejszej części pracy, jednostka ta charakteryzuje się szerokim zakresem funkcji komunikacyjnych, co sprawia, że może być używana w rozmaitych sytuacjach dialogowych i realizować zgoła odmienne cele pragmatyczne. W animowanym filmie *Shrek* zarejestrowano trzy wystąpienia tej jednostki, z których każde pojawia się w innym kontekście i pełni odmienną funkcję – różnice te mają istotne znaczenie zarówno dla analizy pragmatycznej, jak i dla oceny przyjętych strategii tłumaczeniowych. Dla zilustrowania omawianej kwestii, poniżej ponownie przywołano konkretne przykłady, które będą stanowić podstawę dla dalszej analizy:

1.	44:54	Daj spokój	Эй, небойся	Klid, nebojte se
2.	45:38	Daj spokój, żartowałem	Уймись, я пошутил	No tak, dělám si srandu
3.	55:12	Daj spokój	Что, спятил?	Neblazní

Na wstępie należy zaznaczyć, że w języku rosyjskim i czeskim brak jest jednej, konkretnej jednostki leksykalnej, która mogłaby w pełni oddać wszystkie znaczenia oraz funkcje pragmatyczne realizowane przez polskie wyrażenie *daj spokój*. Z tego względu jego przekład każdorazowo wymaga uwzględnienia kontekstu sytuacyjnego, intencji nadawcy oraz funkcji komunikacyjnej danej wypowiedzi, co skutkuje koniecznością doboru różnych odpowiedników w zależności od konkretnego użycia. Jednostka ta, ze względu na swoją wysoką polifunkcyjność, bywa pomijana w opracowaniach leksykografii jedno- i dwujęzycznej, co dodatkowo utrudnia jej systemowe ujęcie. Tym większą wartość poznawczą stanowi wskazany materiał filmowy, który – dzięki wielokrotnemu użyciu analizowanej jednostki w odmiennych kontekstach – pozwala wskazać potencjalne odpowiedniki funkcjonalne, realizujące poszczególne funkcje wyrażenia źródłowego. W pierwszym z zarejestrowanych przypadków, wyrażenie *daj spokój* pojawia się w sytuacji, w której bohater

stara się uspokoić rozmówców – osła oraz księżniczkę – i pełni funkcję fatyczno-regulującą, stanowiąc apel o opanowanie emocji. Przekłady zastosowane w wersjach rosyjskiej i czeskiej – odpowiednio: *Эй, не бойся* oraz *Klid, nebojte se* – zachowują funkcję komunikacyjną oryginału, choć różnią się zarówno pod względem struktury gramatycznej, jak i stopnia bezpośredniości oraz obrazowości wyrażenia. W celu weryfikacji poprawności zastosowania wskazanych przekładów dla polskiego wyrażenia *daj spokój* w kontekście wyrażania potrzeby uspokojenia rozmówcy oraz neutralizacji jego obaw, przeprowadzona zostanie kwerenda korpusowa, uwzględniająca dane z polsko-rosyjskiej wersji korpusu InterCorp v16ud, równoległego korpusu polsko-rosyjskiego NKJR, a także polsko-czeskiej wersji InterCorp v16ud. Cel analizy nie sprowadza się do odnalezienia identycznych realizacji leksykalnych (jak *Эй, не бойся* czy *Klid, nebojte se*), lecz form zbliżonych semantycznie i funkcjonalnie, realizujących ten sam cel komunikacyjny. W tym kontekście uwzględnione zostaną warianty takie jak *не бой*; *не бойся*; *neboj*; *neboj se* itp., które lepiej odzwierciedlają rzeczywiste użycie w przekładach sytuacyjnych.

Jednostka *daj spokój*, wyszukana w wyżej wymienionych korpusach, posiada dużą liczbę poświadczeń: 14 145 w przekładach polsko-rosyjskich i 29 111 w polsko-czeskich, co potwierdza jej wysoką frekwencję i funkcjonalną zmienność. Ze względu na rozpiętość kontekstów i wielość funkcji, analizie poddano wyłącznie te przypadki, które odpowiadają wskazanej funkcji, tj. służą uspokojeniu rozmówcy, zgodnie z sytuacją przedstawioną w filmie. Poniżej zaprezentowano wybrane konkordancje ilustrujące funkcjonalnie równoważne realizacje omawianej jednostki w praktyce tłumaczeniowej:

- Potrzebuję podpisu. Nie twojego. Admirala Rushbrooke'a. Staruszek jest zbyt zajęty. - Nie mogę . Mogę stracić pracę. - <u>Daj spokój.</u> Monday. Zaszalej. Na twoją odpowiedzialność.	- Мне нужна подпись. Нет-нет, не ваша. Адмирала Раишбрука. Старик слишком занят. - Вы знаете что я не могу . Это будет стоить мне работы. -Ну же, Мандэй. <u>Не бойтесь.</u> Пусть это будет на вашей совести .
- Lynette, jesteś pewna że możesz wracać do pracy? - Pewnie! <u>Daj spokój.</u> Wszystko w porządku.	- Ты уверена, что готова вернуться? – Да, <u>не бойтесь.</u> Я в порядке.
- Co to do cholery jest? To jakiś zbiornik. Coś jest w tym dymie. - Podejdź bliżej Clem, chodź zobaczyc. <u>Daj spokój.</u> Nie bój się.	- Что за чертовщина? Что-то вроде резервуара. Там что-то в дыме. - Клем, иди взгляни. О, перестань. <u>Не бойся.</u>
- Mamo, dlaczego tata boi się robaków a ty nie? - Nie boję się wszystkich robaków, tylko... Uważajcie na zardzewiałe gwoździe. - Tato , <u>daj spokój.</u>	- Mami, jakto , že se nebojiš brouků jako táta? - Já se nebojim žádných brouků. Jenom... Pozor na rezavé hřebíky. - <u>Neboj se</u>

-Mam co do tego okrutnie złe przeczucia. - <u>Daj spokój</u> , poradzimy sobie.	-Jimá mne zlověstné, neblahé tušení. - <u>Ale</u> <u>neboj</u> , nějak to přežijeme.
- Czekaj, chyba coś słyszałam. - <u>Daj spokój</u> , maleńka. - Nie ma tu nikogo poza tobą i mną.	-Počkej. Něco jsem slyšela. - Ale jdi, <u>neboj</u> <u>se</u> , drahá. Jsme tady jenom my dva.

Przeprowadzona analiza potwierdza zasadność tłumaczenia polskiego wyrażenia *daj spokój* – w kontekście jego funkcji uspokajającej – za pomocą takich jednostek jak rosyjskie *не бойся* 'не бойтесь' czy czeskie *neboj*; *neboj se* itp. Pomimo faktu, iż wskazane formy różnią się strukturalnie od jednostki wyjściowej, są wobec niego funkcjonalnie równoważne – realizują ten sam cel komunikacyjny: apel o opanowanie emocji i redukcję napięcia. Zastosowane przekłady mają charakter opisowy i sytuacyjny, co wskazuje na wykorzystanie strategii ekwiwalencji funkcjonalnej, polegającej na doborze takich środków językowych, które umożliwiają zachowanie funkcji wypowiedzi w danym kontekście. Poświadczenia korpusowe potwierdzają zatem zasadność tłumaczenia *daj spokój* w analizowanym znaczeniu za pomocą wskazanych jednostek, zgodnie z obserwacjami wynikającymi z materiału filmowego.

Przejdźmy do analizy drugiego zarejestrowanego użycia wyrażenia *daj spokój*, które – choć nadal pełni funkcję uspokajającą – zostało wzmocnione dodatkową frazą *żartowałem*, stanowiącą element usprawiedliwienia wcześniejszej wypowiedzi bądź zachowania nadawcy. W danym przykładzie wyrażenie pełni funkcję regulującą i służy wprowadzeniu dystansu i podkreśleniu żartobliwego charakteru sytuacji. W wersji rosyjskiej zastosowano wyrażenie *Уймись, я поушмил*, w którym pierwsza część – imperatyw *Уймись* ('uspokój się') – w sposób bezpośredni apeluje o opanowanie emocji rozmówcy, natomiast fraza *я поушмил* („żartowałem”) pełni funkcję wyjaśniająco-usprawiedliwiającą. Choć strukturalnie przekład ten jest bardziej rozwinięty niż jednostka wyjściowa, zachowuje jej zasadniczą funkcję komunikacyjną.

Z kolei przekład czeski przyjmuje postać *No tak, dělám si srandu*, który również spełnia funkcję pragmatycznie równoważną. Na szczególną uwagę zasługuje tutaj obecność wprowadzenia *no tak*, które w języku czeskim funkcjonuje jako metaoperator dyskursywny. W analizowanym przypadku *no tak* wskazuje na łagodne wycofanie się z wcześniejszej wypowiedzi, nadaje jej ton potoczny oraz autoironiczny. W danym ujęciu reprodukt *No tak, dělám si srandu* można zatem interpretować jako *No przecież żartuję*. Zarówno przekład rosyjski, jak i czeski realizują funkcjonalną równoważność wobec jednostki źródłowej, zachowując jej intencję i efekt komunikacyjny pomimo widocznych różnic strukturalnych. W oparciu o te same korpusy równoległe oraz z wykorzystaniem tożsamej listy konkordancji, poniżej przedstawiono przykłady poświadczeń, w których jednostka *daj spokój* została

przetłumaczona na język rosyjski i czeski w sposób funkcjonalnie zbliżony do analizowanego przypadku:

- Nie pomaga, że jesteś w tym beznadziejna. Mówię poważnie... <u>Daj spokój</u> , żartuję.	- Но то, что ты в этом плоха, делу не поможет. Да, брось, <u>я же пошутил</u> .
- Niezła próba, ale ta dziewczyna ma wysokie standardy. Wiem to z pierwszej ręki... <u>Daj spokój</u> , to był żart. - Nieśmieszne.	- Попытка засчитана, но у нее более высокие требования. Кому, как не мне, знать... Ну же, <u>я пошутил</u> . – Несмешно.
- Del, ty jelo pie. Nie miałem złych zamiarów. Wyglądałem się. <u>Daj spokój</u> .	- Дел, ты просто дурак. Я это сделал просто так. <u>Я только пошутил</u> .
- Nie mam chłopaka. A ty ? - Nie, jestem lesbijką... <u>Daj spokój</u> , żartuję.	- Ne, nemám žádného přítele. A co ty? - Ne, sakra. Já jsem lesbička... <u>Dělám si srandu</u> , vždyť mě znáš.
- To jak żywcem wyjęte z kinowych horrorów. Te stworzenia powstają z martwych. <u>Daj spokój</u> , żartowałem.	- To je jako z nějakého hororu. Stvoří co vylézají z hrobů... Tak jo, <u>dělám si srandu</u> .
- <u>Daj spokój</u> , przecież żartuję! Przecież mnie znasz. Jestem spoko ojcem.	- <u>Dělám si srandu</u> ! No tak, za koho mě máš? Jsem pohodovej táta.

Wybrane poświadczenia korpusowe potwierdzają, że wyrażenie *daj spokój* w funkcji usprawiedliwienia żartobliwej wypowiedzi lub zachowania jest tłumaczone w sposób zgodny z tendencją zaobserwowaną w materiale filmowym. Choć struktury *Уймишь, я пошутил* oraz *No tak, dělám si srandu* nie występują w analizowanych przykładach w identycznym układzie, to jednak ich poszczególne elementy – zarówno komunikaty o charakterze pragmatycznym (np. *я пошутил*; *dělám si srandu*), jak i środki o funkcji regulującej (np. *Да брось*; *Ну же*; *Tak jo*) – są obecne i pełnią analogiczną rolę. Zarówno w rosyjskich, jak i czeskich przekładach dostrzegalna jest strategia funkcjonalnego dopasowania, w ramach której tłumacze wykorzystują różne komponenty językowe do osiągnięcia tej samej funkcji: złagodzenia skutków wypowiedzi, wycofania jej dosłownej interpretacji oraz utrzymania kontaktu z odbiorcą. Warto przy tym zwrócić uwagę na obecność metaoperatorów dyskursywnych, takich jak rosyjskie *Да брось*; *Ну же* czy czeskie *Tak jo*, które pełnią funkcję organizującą wypowiedź, sygnalizują dystans lub pozwalają nadawcy przejść do wyjaśnienia.

Zgromadzone poświadczenia potwierdzają zatem, że w analizowanej funkcji *daj spokój* jest tłumaczone podobnie jak w filmie, choć należy podkreślić, iż to tylko jedna z możliwych strategii przekładu. Ostatni z analizowanych przykładów przedstawia użycie wyrażenia *daj spokój* jako reakcji wyrażającej dezaprobatę, odrzucenie lub zaskoczenie wobec wypowiedzi bądź zachowania rozmówcy. W tym kontekście jednostka ta pełni funkcję ekspresywną, sygnalizując emocjonalny sprzeciw i brak akceptacji wobec powstałej sytuacji. W przekładzie rosyjskim zastosowano frazę *Что, снятил?*, w której element *что* nie pełni typowej funkcji pytającej, lecz służy jako środek wprowadzający wypowiedź ekspresywną, charakterystyczną dla sytuacji nacechowanych emocjonalnie. W połączeniu z czasownikiem *снятил* konstrukcja ta wyraża silny sprzeciw i niedowierzanie, odpowiadając znaczeniowo potocznym polskim reakcjom typu *Co ty, zwariowałeś?* w funkcji ekspresywnej dezaprobaty. Z kolei przekład czeski przyjmuje formę jednowyrazowego *Neblázni*, które – mimo swojej zwięzłości – pełni funkcję ekspresywną zbliżoną do polskiego *daj spokój, nie wygłupiaj się*. Wyrażenie to, typowe dla języka potocznego, służy odrzuceniu lub zakwestionowaniu wypowiedzi rozmówcy, wyrażając zaskoczenie lub dezaprobatę wobec jej treści. Na podstawie tych samych zasobów korpusowych i zestawu konkordancji zaprezentowano poniżej kolejne przykłady tłumaczeń jednostki *daj spokój*:

- Nawet ze mną nie porozmawiasz? Dzwoniłam dzisiaj 8 razy. – No, <u>daj spokój</u> . Miałam zły nastrój.	- Ty se mnou nebudeš mluvit? Dnes jsem ti volala osmkrát. - <u>Neblázni</u> . Měla jsem blbou náladu.
- Miriam! Miriam, gdzie jesteś? Miriam! Miriam, <u>daj spokój</u> . Przestań się wygłupiać. Wydaje ci się, że nam uciekłaś, co ?	- Miriam! Miriam, kde jsi? Miriam! Miriam, <u>neblázni</u> . Chceš nám zdrhnout?
- Hej, Spider. Daj mi pozmywać, to mi robi dobrze. - <u>Daj spokój</u> , jesteś moim gościem.	- Hej, Pavouku! Nech mě umýt nádobí. Je mi líp. - <u>Neblázni</u> , jsi můj host.

Jak można zauważyć, w poświadczeniach korpusowych nie odnotowano jednostki *Что, снятил?* jako tłumaczenia wyrażenia *daj spokój*, co może wskazywać na jej ograniczone zastosowanie w praktyce przekładowej bądź wynikać z charakteru samych korpusów, które rzadko rejestrują ekspresywne jednostki potoczne. W odróżnieniu od tego, w wersjach czeskich odpowiednik *neblázni* pojawia się wielokrotnie jako ekwiwalent w funkcji dezaprobaty lub zaskoczenia, co potwierdzają wskazane poświadczenia z korpusów. Świadczy to o ugruntowanej pozycji tego wyrażenia w czeskim języku potocznym jako środka realizującego funkcję ekspresywnego sprzeciwu wobec wypowiedzi rozmówcy. W porównaniu z

przekładami czeskimi, które zachowują wysoką zgodność funkcjonalną i są reprezentatywne dla tej funkcji, przekład rosyjski z filmu ma charakter bardziej idiomatyczny i nacechowany emocjonalnie. W związku z brakiem poświadczeń korpusowych dla jednostki *Что, спятил?* jako tłumaczenia translandu polskiego, zasadne jest odwołanie się do źródeł pozakorpusowych w celu określenia jej znaczenia i poprawności użycia.

Большой толковый словарь русского языка pod redakcją S. A. Kuzniecowa opisuje czasownik *спатить* jako wyraz potoczny, oznaczający 'zwariować', 'oszaleć' (*сойти с ума; помешаться*). Artykuł hasłowy wskazuje na ekspresywno-emocjonalny charakter jednostki, czego przykładem jest użycie *Ты что, спятил?* – tożsame z konstrukcją zastosowaną w filmie. Potwierdza to zatem jej funkcję jako środka ekspresywnego sprzeciwu typowego dla języka potocznego (Кузнецов 2000: 1254). Dodatkowe źródła, w tym wypowiedzi użytkowników na forach internetowych, potwierdzają przedstawioną interpretację analizowanej jednostki. Można założyć, iż konstrukcja ta jest znaczeniowo i strukturalnie zbliżona do wyrażenia *oszalałeś?*, jednak w określonych kontekstach może pełnić funkcję ekwiwalentu funkcjonalnego dla polskiego *daj spokój*. Zgromadzony materiał potwierdza zatem zasadność przyjętych przekładów w obu wersjach językowych, które realizują funkcję ekspresywnego sprzeciwu, stanowiąc sytuacyjne ekwiwalenty emotywu *daj spokój*.

Można zatem podsumować, że zastosowanie metody audiowizualnej posiada duży potencjał badawczy, szczególnie przy identyfikacji funkcjonalnych odpowiedników w kontekstach dialogowych. Analiza dubbingowanego filmu animowanego pozwoliła na wyodrębnienie licznych trójjęzycznych zestawień przekładowych (polsko–rosyjsko–czeskich), z których część – mimo różnic strukturalnych – wykazuje zgodność pragmatyczną i może być traktowana jako sytuacyjnie adekwatne odpowiedniki tłumaczeniowe. Szczególnie cenne są te jednostki, które zachowują spójność funkcji we wszystkich trzech wersjach językowych i osadzone są w jasno określonym kontekście sytuacyjnym. Uzyskane translaty nie mają jednak charakteru pełnych ekwiwalentów leksykalnych. Są to rozwiązania wynikające z konkretnych decyzji translatorskich, podejmowanych w warunkach przekładu filmowego, który – z uwagi na ograniczenia czasowo-przestrzenne i konieczność dopasowania do obrazu – dopuszcza większą swobodę i adaptację wypowiedzi. Mimo tego, mogą one stanowić wartościowy punkt wyjścia do dalszych badań porównawczych oraz selekcji jednostek przeznaczonych do słowników przekładowych, gdyż ich obecność w dubbingu sugeruje ich praktyczne użycie w komunikacji codziennej. Zaletą metody tej jest również możliwość uchwycenia jednostek potocznych i pragmatycznie nacechowanych, które są słabo reprezentowane w tekstach pisanych i korpusach. Ze względu na sytuacyjny charakter danych oraz ograniczoną

reprezentatywność analizowanych tłumaczeń, metoda ta nie może jednak pełnić samodzielnej funkcji opisowej, lecz powinna być traktowana jako narzędzie pomocnicze. Dostarcza ona autentycznego materiału ilustracyjnego, wspiera identyfikację ekwiwalentów funkcjonalnych i może być wykorzystywana jako uzupełnienie analiz korpusowych przy opracowywaniu słowników jedno- i wielojęzycznych oraz w badaniach nad przekładalnością języka mówionego.

7.3 Metoda semantyczno-pragmatyczna

Ostatnią z zaprezentowanych technik dochodzenia do ekwiwalentów dla fatyzmów i emotywów stanowić będzie tzw. *metoda semantyczno-pragmatyczna*, oparta na analizie wcześniej opracowanej bazy wielowyrazowych jednostek języka polskiego, rosyjskiego i czeskiego. Jak już wielokrotnie wspomniano, baza ta została zbudowana w oparciu o źródła korpusowe i pozakorpusowe, a dla każdej z jednostek określono jej semantykę oraz funkcję pragmatyczną, z uwzględnieniem typowych użyc w sytuacji komunikacyjnej. Następnie wszystkie jednostki zostały pogrupowane według kategorii znaczeniowo-funkcjonalnych, wspólnych dla trzech badanych języków.

Metoda semantyczno-pragmatyczna różni się od wcześniej omówionych metod przede wszystkim punktem wyjścia. W tym przypadku analiza nie koncentruje się na wyjściowo wybranej jednostce leksykalnej w jednym języku, lecz odwołuje się do wspólnego znaczenia lub funkcji, ustalonych wcześniej na etapie klasyfikacji materiału. W procesie budowania bazy produktów zaobserwowano, że pomimo różnic formalnych (strukturalnych, leksykalnych, obrazowych), niektóre jednostki w trzech badanych językach pełnią analogiczne funkcje pragmatyczne i są używane w analogicznych sytuacjach komunikacyjnych. Należy podkreślić, iż jednostki te funkcjonowały w zbiorze jako samodzielne elementy przypisane do poszczególnych języków, a nie jako składniki par przekładowych. Dopiero ich zestawienie w obrębie wspólnej kategorii semantyczno-pragmatycznej ujawniło potencjał do ustanowienia konkretnych par przekładowych opartych na zgodności znaczeniowo-funkcjonalnej.

Metoda polega zatem na identyfikacji takiego znaczenia, które występuje równolegle w bazie jednostek polskich, rosyjskich i czeskich, a następnie na wyodrębnieniu wszystkich form przyporządkowanych do danej kategorii w każdym z języków. Na tym etapie nie istnieje potrzeba ponownej weryfikacji znaczenia poszczególnych jednostek w źródłach zewnętrznych, ponieważ przypisane im funkcje oraz definicje zostały już wcześniej ustalone i udokumentowane na etapie budowy bazy. Celem obecnej procedury pozostaje zatem

utworzenie powiązań przekładowych pomiędzy jednostkami z trzech języków, które posiadają podobne bądź takie samo znaczenie. Kluczowym etapem jest wewnętrzna analiza porównawcza, polegająca na zestawieniu pod względem formalnym i znaczeniowym form najbardziej zbliżonych. Należy podkreślić, iż liczba jednostek przyporządkowanych do określonej funkcji w poszczególnych językach nie zawsze jest jednakowa. W takich przypadkach jednostki liczniej reprezentowane w jednym z języków, a niepołączone w pary przekładowe w obrębie zebranego zbioru, mogą stanowić punkt wyjścia do poszukiwania brakujących odpowiedników w pozostałych. Takie podejście sprzyja również samemu poszerzaniu bazy wielowyrazowych fatyzmów i emotywów. Końcowy etap metody stanowić będzie weryfikacja zaproponowanych par przekładowych głównie w korpusach równoległych (np. InterCorp, NKJR). Choć celem nie jest statystyczna analiza frekwencji, dane te pozwalają sprawdzić, czy ustalone zestawienia są zasadne z punktu widzenia rzeczywistej praktyki tłumaczeniowej. W sytuacji, gdy zaproponowane pary nie znajdą odzwierciedlenia w danych korpusowych, możliwa jest ich klasyfikacja jako ekwiwalentów zbliżonych, wykazujących zgodność semantyczno-pragmatyczną, lecz niekoniecznie potwierdzoną w tekstach. Na potrzeby przeprowadzenia analizy oraz w celach egzemplifikacyjnych wyodrębniono z bazy wielowyrazowych jednostek językowych grupę emotywów reakcyjnych, które w strukturze wypowiedzi występują zazwyczaj na jej początku lub końcu. Jednostki te realizują znaczenie polegające na jednoznacznym zakomunikowaniu, iż decyzja mówiącego ma charakter ostateczny, nie podlega dyskusji i wymaga natychmiastowego oraz bezwarunkowego podporządkowania się odbiorcy. W tabeli poniżej przedstawiono zestawienie polskich, rosyjskich i czeskich emotywów przyporządkowanych tej funkcji wraz z odpowiadającymi im definicjami:

ZNACZENIE		
Używane w celu wyraźnego zaznaczenia, że decyzja mówiącego jest nieodwołalna i wymaga natychmiastowego podporządkowania się bez prób jej kwestionowania.	Употребляется для выражения категорического запрета на возражения и обсуждения, обозначая требование безусловного подчинения и выполнения чего-либо без дополнительных условий.	Používá se k jasnému vyjádření, že postoj mluvčího je konečný a nepřipouští žádnou diskusi, a proto se od adresáta očekává okamžité a bezvýhradné podřízení.
Emotywy polskie	Emotywy rosyjskie	Emotywy czeskie
Żadnych ale!	Никаких но!	Žádné ale!
To moje ostatnie słowo!	Это моё последнее слово!	To je moje poslední slovo!
Bez dyskusji!	Без разговоров!	Žádné výmluvy!
Bez gadania!	Разговор окончен!	Bez řečí!
	Больше ни слова!	

Jak można zauważyć, w obrębie wybranej grupy znaczeniowej największą liczbę zarejestrowanych jednostek odnotowano w języku rosyjskim, w którym występuje pięć wariantów realizujących wskazaną funkcję. Wstępna obserwacja materiału ujawnia również wyraźne podobieństwa formalne pomiędzy niektórymi jednostkami w trzech językach, szczególnie w przypadku polskiego

Żadnych ale! – rosyjskiego *Никаких но!* – czeskiego *Žádné ale!*, a także

To moje ostatnie słowo! – *Это моё последнее слово!* – *To je moje poslední slovo!*

Podobieństwa te dotyczą zarówno struktury składniowej, jak i doboru komponentów leksykalnych, co pozwala uznać je za najbardziej zbliżone, pełne ekwiwalenty. Powyższe zestawienia, ze względu na swoją bezpośrednią przekładalność, zostały przyjęte jako ustalone odpowiedniki i w toku dalszej analizy nie będą już szczegółowo omawiane. Jednocześnie w tabeli obecne są jednostki, które pomimo realizowania tej samej funkcji, zauważalnie różnią się strukturą i obrazowością, co stanowi potencjalny materiał do dalszej analizy kontrastywnej. Należy zatem przeanalizować, które z jednostek ujętych w zestawieniu stanowią względem siebie najbardziej zbliżone przekłady, czy też posiadają inny, częściej występujący ekwiwalent, co wymaga odwołania się do źródeł korpusowych. W tym celu każda z jednostek (polska, rosyjska i czeska) zostanie wyszukana w korpusie InterCorp v16ud oraz w korpusie równoległym NKJR, aby na podstawie poświadczeń tekstowych ustalić najczęściej stosowane dla niej ekwiwalenty. Dla każdego z wyszukanych translatów zostanie określony jego udział procentowy w ogólnej liczbie poświadczeń, a wyniki zestawiono w tabelach przedstawionych poniżej:

Tabela nr 1

Jednostka polska	Ekwiwalenty rosyjskie	Ekwiwalenty czeskie
Bez dyskusji!	Это не обсуждается! (14,2%)	Bez diskuse! (11,7%)*
	Не спорь! (5,2%)	Bez debat! (2,4 %)
		Bez odmlouvání! (1,3%)
Bez gadania!	Никаких разговоров! (4,6%)	Bez řečí! (4,6%)
	Без возражений! (3,7%)	Žádné výmluvy! (2,9%)
	Без разговоров! (2,7%)	

Tabela nr 2

Jednostka rosyjska	Ekwiwalenty polskie	Ekwiwalenty czeskie
Без разговоров!	Bez gadania! (8,6%)	Bez řečí! (5,3%)
	Nie gadać! (5,4%)	Žádné mluvení! (4,4%)
Разговор окончен!	Koniec rozmowy (14,9%)	Koniec diskuse (8,6%)
	Koniec dyskusji (5,9%)	Rozhovor skončil (8%)
Больше ни слова!	Ani słowa! (8,9%)*	Už ani slovo!* (13,1%)
	Nic nie mów! (2,1%)	Už nic neříkej/te! (10%)

Tabela nr 3

Jednostka czeska	Ekwiwalenty polskie	Ekwiwalenty rosyjskie
Žádné výmluvy!	Žadnych wymówek! (20,3%)	Никаких отговорок! (19,3%)
	Bez wymówek! (5,3%)	Никаких оправданий! (14,6%)
Bez řečí!	Bez gadania! (5%)	Без разговоров! (4,7%)
	Cisza! (1%)	Без возражений! (2,8%)

Pierwsza z analizowanych jednostek, tj. *Bez dyskusji!* w tłumaczeniach korpusowych najczęściej odpowiada rosyjskiej frazie *Это не обсуждается!* (14,2%) oraz czeskiej formie *Bez diskuse!* (11,7%). Należy jednak zauważyć, że rosyjski odpowiednik jest prawdopodobnie pod względem semantycznym bliższy polskiemu wyrażeniu *To nie podlega dyskusji*. W przypadku poświadczeń czeskich wskazany udział procentowy dla najczęściej używanego translatu uwzględnia również warianty *bez diskuze*, *bez diskuś* oraz *bez diskuží*, przy czym formy *diskuse* oraz *diskuze* to formy oboczne. W obu językach odnotowano również inne przykłady ekwiwalentów, m.in. rosyjskie *He cnoпb!* (5,2%) oraz czeskie *Bez debat!* (2,4%) i *Bez odmłouvání!* (1,3%), które choć mogą poszerzyć bazę jednostek wielowyrazowych do dalszej analizy, na obecnym etapie nie są uwzględniane jako bezpośrednie odpowiedniki badanej jednostki. Na podstawie uzyskanych wyników można zatem zaproponować następującą trójjęzyczną triadę przekładową:

Bez dyskusji! – Это не обсуждается! – Bez diskuse!

Z kolei omówienie jednostki *Bez gadania!* wymaga odniesienia do wyników przedstawionych w kilku tabelach, ponieważ występuje ona zarówno jako jednostka wyjściowa, jak i jako najczęstszy odpowiednik w tłumaczeniach jednostek rosyjskich i czeskich.

W przekładzie polsko-rosyjskim (tab. 1) najwyższy odsetek poświadczeń uzyskała forma *Никаких разговоров!* (4,6%), przy czym w zestawieniu znalazła się również forma *Без разговоров!*, która – w kierunku rosyjsko-polskim (tab. 2) – była tłumaczona właśnie jako *Bez gadania!*.

Natomiast w przekładzie polsko-czeskim (tab. 1) najczęstszym odpowiednikiem analizowanej jednostki jest *Bez řečí!*, która w odwrotnym kierunku (czesko-polskim, tab. 3) również występuje jako ekwiwalent główny dla *Bez gadania!*. Obie te jednostki – *Без разговоров!* w rosyjskim i *Bez řečí!* w czeskim – pojawiają się także jako najczęstsze translaty w przekładach czesko-rosyjskich (tab. 3). Uwzględniając zbieżność frekwencyjną i funkcjonalną, za adekwatną trójjęzyczną triadę przekładową można uznać zatem potoczne

Bez gadania! – Без разговоров! – Bez řečí!.

W ramach zestawień tłumaczeń rosyjsko-polskich oraz rosyjsko-czeskich, po pominięciu wcześniej opisanych jednostek, przeanalizowano frazę *Разговор окончен!*, której przekłady na język polski i czeski mogą budzić wątpliwości z uwagi na ich frekwencję. W poświadczeniach korpusowych w przekładzie na język polski najczęściej występuje forma *koniec rozmowy* (14,9%), będąca efektem dosłownego tłumaczenia rosyjskiego translandu. Choć ma ona zbliżone znaczenie pragmatyczne do analizowanej jednostki, obserwacja przykładów z NKJP wskazuje, że w polszczyźnie pojawia się rzadziej w kontekstach odpowiadających tej funkcji niż *koniec dyskusji* (5,9%), typowe dla sytuacji wymagających kategorycznego zakończenia rozmowy i bezwarunkowego podporządkowania się.

W części czeskiej korpusu InterCorp v16ud większą odpowiedniość formy *konec diskuse* potwierdza analiza kontekstów, które wskazują na jej częstsze użycie w znaczeniu zgodnym z przyjętą definicją. Z tego względu, mimo niższej frekwencji w materiale polskim, w dalszej analizie przyjęto formy *koniec dyskusji* oraz *konec diskuse* jako najbardziej zbliżone. W dalszej części tabeli nr 2 poddano jednostkę *Больше ни слова!*, która w przekładzie na język polski najczęściej oddawana była formą *Ani słowa!* (8,9%). Wskazany udział procentowy uwzględnia również wariant *Ani słowa więcej!*, będący semantycznie bardzo zbliżoną realizacją danego reproduktu.

W tłumaczeniach na język czeski dominującym odpowiednikiem okazało się wyrażenie *Už ani slovo!* (13,1%), przy czym dane frekwencyjne uwzględniały także skrócone formy typu

Ani slovo!. Pozostałe poświadczenia, takie jak polskie *Nic nie mów!* czy czeskie *Už nic neříkej/te!*, pełnią zbliżone funkcje i odznaczają się znacznym podobieństwem formalnym, jednak na obecnym etapie analizy nie zostaną włączone do proponowanej triady przekładowej. Na podstawie przedstawionych danych za trójjęzyczny ciąg ekwiwalentów można uznać:

Больше ни слова! – Ani słowa! (Ani słowa więcej!) – (Už) ani slovo!.

W tabeli nr 3, w części poświęconej tłumaczeniom czesko-polskim i czesko-rosyjskim, szczegółowej analizie poddano jednostkę *Žádné výmluvy!*, natomiast ekwiwalenty dla *Bez řečí!* zostały omówione we wcześniejszej części analizy. W przekładzie na język polski najczęściej stosowanym odpowiednikiem jest forma *Żadnych wymówek!* (20,3%), natomiast w przekładzie na język rosyjski – *Никаких отговорок!* (19,3%). Oba te translaty można uznać za bezpośrednie kalki, cechujące się wysokim stopniem zgodności semantycznej i strukturalnej z jednostką wyjściową. Wśród poświadczeń występują również inne warianty, takie jak polskie *Bez wymówek!* (5,3%) oraz rosyjskie *Никаких оправданий!* (14,6%), które mimo częściowych różnic leksykalnych, zachowują tożsamy sens komunikacyjny, stanowiąc warianty ekwiwalentów głównych. Uwzględniając powyższe dane, za trójjęzyczną triadę przekładową można przyjąć:

Žádné výmluvy! – Žadnych wymówek! – Никаких отговорок!.

Przeprowadzona analiza pozwoliła na wskazanie następujących trójjęzycznych zestawień przekładowych, uznanych za adekwatne pod względem zgodności semantyczno-pragmatycznej:

Bez dyskusji! – Это не обсуждается! – Bez diskuse!;

Bez gadania! – Без разговоров! – Bez řečí!;

Разговор окончен! – Koniec dyskusji! – Konec diskuse!;

Больше ни слова! – Ani słowa! (Ani słowa więcej!) – (Už) ani slovo!;

Žádné výmluvy! – Žadnych wymówek! – Никаких отговорок!.

Powyższa propozycja stanowi wynik procesu porównawczej analizy formalnej i funkcjonalnej, prowadzonej na podstawie danych korpusowych i obserwacji kontekstów użycia. Należy jednak podkreślić, iż granice pomiędzy wskazanymi ekwiwalentami są płynne, a ich zakresy użycia mogą się przenikać. Wybrane jednostki uznano za najbardziej zbliżone pod względem funkcji i struktury, lecz nie wyczerpują one wszystkich możliwych rozwiązań przekładowych. Badanie wykazało również rozbieżności między danymi frekwencyjnymi a oceną adekwatności, np. *koniec rozmowy* czy *rozhovor skončil* są częste w korpusach, lecz rzadziej spełniają funkcję kategorięcznego zakończenia dyskusji niż *koniec dyskusji* i *konec*

diskuse. Pokazuje to konieczność łączenia kryteriów ilościowych z analizą jakościową kontekstów użycia.

Podsumowując, celem przeprowadzonej procedury było ustalenie takich par przekładowych, które w możliwie największym stopniu zachowują spójność znaczeniowo-funkcjonalną w trzech badanych językach. Powstałe zestawienie należy traktować jako propozycję najbardziej zbliżonych odpowiedników w obrębie analizowanej funkcji, przy jednoczesnym uwzględnieniu, że w praktyce przekładowej mogą być stosowane także inne warianty, zależnie od kontekstu oraz intencji nadawcy.

W tym miejscu przejdźmy do analizy fatyzmów, która będzie przeprowadzona w sposób analogiczny do wcześniejszego opisu grupy emotywów reakcyjnych. Przedmiotem tej części są jednostki realizujące funkcję wprowadzania dodatkowych treści do wypowiedzi. Zebrane jednostki służą podkreśleniu, że przekazywana informacja stanowi jedynie fragment całego komunikatu i wymaga uzupełnienia o kolejne elementy, które poszerzają lub wzmacniają jego treść. Analiza poświadczeń korpusowych, przeprowadzona na etapie budowy bazy, potwierdziła, że jednostki te najczęściej występują na początku lub w środkowej części wypowiedzi. W tabeli poniżej zaprezentowano zestawienie polskich, rosyjskich i czeskich fatyzmów przyporządkowanych tej funkcji wraz z odpowiadającymi im definicjami:

ZNACZENIE		
Używane w celu zaakcentowania, że przekazywana informacja stanowi jedynie część wypowiedzi i wymaga uzupełnienia o kolejne elementy, które poszerzają lub wzmacniają jej treść	Употребляется для подчёркивания, что сказанное является лишь частью сообщения и вводит новые элементы, усиливающие его; подготавливает к более сильному заключительному утверждению.	Používá se k zdůraznění, že sdělené představuje pouze část informace, a taky k přidání nových prvků, které ji rozšiřují nebo posilují; připravuje na silnější závěrečné tvrzení
Fatyzmy polskie	Fatyzmy rosyjskie	Fatyzmy czeskie
A w dodatku...	Ко всему прочему...	To není všechno...
A do tego...	К тому же...	A k tomu...
Mało tego...	Кроме того...	A navíc...
A co więcej...	Да ещё...	Kromě toho...

Analiza danych przedstawionych w tabeli wskazuje, że ujęte w niej jednostki, oprócz potwierdzonej wcześniej bliskości znaczeniowej, posiadają także zbliżoną strukturę, co pozwala na ich stosowanie w funkcji synonimicznej. Obserwacja materiału pozwala również od razu dostrzec wysoki stopień zgodności formalno-strukturalnej pomiędzy niektórymi jednostkami, czego przykładem jest zestawienie

A do tego... – К тому же... – A k tomu....

Zwraca również uwagę obecność rosyjskiej formy *Кrome mozo...* i czeskiej *Kromě toho...*, dla których brak bezpośredniego odpowiednika polskiego w tabeli, jednakże już na danym etapie można założyć, iż właściwym odpowiednikiem byłby rezydent *Oprócz tego*. Pomimo ogólnej ekwiwalencji funkcjonalnej i tego samego znaczenia, formy te różnią się stopniem zbieżności formalnej. W związku z tym podjęta zostanie próba wyodrębnienia najbliższych formalnie wariantów, co umożliwi ustalenie ekwiwalentów oraz uzupełnienie bazy jednostek fatycznych w tej kategorii. W tym celu, analogicznie do poprzedniej analizy, jednostki pozbawione par zostaną wyszukane w korpusie InterCorp v16ud oraz w korpusie równoległym NKJR. Na podstawie poświadczeń tekstowych zostaną określone ich najczęściej stosowane odpowiedniki, wraz z podaniem udziału procentowego w ogólnej liczbie poświadczeń. Wyniki zestawiono w tabelach poniżej:

Tabela nr 1

Jednostka polska	Ekwiwalenty rosyjskie	Ekwiwalenty czeskie
A w dodatku...	К тому же... (10,1%)	A navíc... (15,5%)
	Кrome того... (7,7%)	Kromě toho... (4,1%)
	И вдобавок... (4,7%)	A ještě k tomu... (2,1%)
Malo tego...	Мало того... (17,5%)	A nejen to... (7%)
	Более того... (5,4%)	Víc než to... (3,5%)
	К тому же... (4%)	A navíc... (3,5%)
A co więcej...	Более того... (22,3%)	A co víc... (16,2%)
	Кrome того... (5,8%)	A navíc... (11,6%)

Tabela nr 2

Jednostka rosyjska	Ekwiwalenty polskie	Ekwiwalenty czeskie
Ко всему прочему...	I do tego (wszystkiego)... (7%)	Ještě ke všemu... (10,1%)
	W dodatku... (5,3%)	A navíc... (8,5%)
Да ещё	I do tego... (3,2%)	A navíc... (11,3%)
	I jeszcze... (2,1%)	A ještě... (9,7%)

Tabela nr 3

Jednostka czeska	Ekwiwalenty polskie	Ekwiwalenty rosyjskie
To není všechno...	To nie wszystko... (31,3%)	Это не всё... (20,4%)
	To nie koniec... (3%)	Есть еще кое-что... (5,1%)
A navíc...	Poza tym... (47,5%)	К тому же... (47,5%)
	A w dodatku... (8,2%)	И вдобавок... (4%)
		Да ещё... (2,7%)

Tabela otwierająca analizę zawiera zestawienia polskich jednostek z ich rosyjskimi i czeskimi odpowiednikami, a jako pierwszy zaprezentowano w niej fatyzm *A w dodatku...*

W materiale rosyjskim najczęściej poświadczonym ekwiwalentem tej jednostki jest *K тому же...* (10,1%), a następnie *Кроме того...* (7,7%).

Oba te wyrażenia są zbliżone znaczeniowo i funkcjonalnie, jednak jak wskazano wcześniej, *K тому же...* odpowiada raczej konstrukcji *A do tego...*, podczas gdy *Кроме того...* jest bliższe polskiemu *Oprócz tego*.

Trzecia odnotowana forma, tj. *И вдобавок...* (4,7%), mimo najniższego udziału procentowego, cechuje się najwyższą zgodnością semantyczną i strukturalną z jednostką polską, co uzasadnia uznanie jej w tym przypadku za najtrafniejszy ekwiwalent.

W przekładach na język czeski najwyższy udział frekwencyjny osiągnęła jednostka *A navíc...* (15,5%). Odnotowano również *Kromě toho...* (4,1%), odpowiadające konstrukcji *Oprócz tego*, oraz *A ještě k tomu...* (2,1%), będące bliższym odpowiednikiem *A jeszcze do tego...*

Zestawiając wyniki dla obu języków, za adekwatne ekwiwalenty jednostki *A w dodatku...* należy uznać zatem rosyjskie *И вдобавок...* oraz czeskie *A navíc...*. Kolejną analizowaną jednostką jest wyrażenie *Mało tego...*, które w materiale rosyjskim najczęściej zostało oddane za pomocą formy *Мало того...* (17,5%). Odpowiednik ten, zarówno pod względem znaczeniowym, jak i strukturalnym, można uznać za pełny ekwiwalent jednostki polskiej, co znajduje potwierdzenie w wysokim udziale frekwencyjnym. W poświadczeniach pojawiły się także inne konstrukcje, takie jak *Более того...* (5,4%) oraz wcześniej omawiane *K тому же...* (4%), które w danym przypadku mogą być stosowane jako synonimy. W przekładach na język czeski procentowo największy udział miało wyrażenie *A nejen to...* (7%). Odnotowano także konstrukcje *Víc než to...* (3,5%) oraz *A navíc...* (3,5%), które choć mogą

występować w zbliżonych kontekstach, stanowią trafniejsze ekwiwalenty innych jednostek niż analizowana polska forma wyjściowa. Biorąc pod uwagę zarówno kryterium frekwencji, jak i stopień zgodności formalnej, za adekwatny trójjęzyczny ciąg przekładowy należy uznać zestawienie *Mało tego... – Мало того... – A nejen to...*

Ostatnią z jednostek ujętych w analizowanej tabeli jest fatyzm *A co więcej...*. W języku rosyjskim najwyższy udział frekwencyjny uzyskało wyrażenie *Более того...* (22,3%), które pod względem semantycznym i strukturalnym stanowi najbliższy odpowiednik formy polskiej. W przekładach na język czeski dominującą pozycję zajmuje konstrukcja *A co víc...* (16,2%), zachowująca podobny poziom zgodności. Uwzględniając te kryteria, za adekwatną triadę przekładową należy uznać *A co więcej... – Более того... – A co víc...*

W tabeli nr 2, poświęconej tłumaczeniom dla wybranych jednostek rosyjskich, pierwszym fatyzmem jest transland *ко всему прочему...*. Wśród poświadczeń korpusowych najwyższą frekwencję w przekładach polskich odnotowano dla ekwiwalentu *i do tego...*, który stanowił 7% wszystkich konkordancji. Należy zaznaczyć, że jednostka ta często występowała w postaci *i do tego wszystkiego...*, stanowiącej możliwą realizację formy inwariantywnej. W przekładach na język czeski najwyższy odsetek uzyskała jednostka *ještě ke všemu...* (10,1%), której znaczenie i struktura są zbliżone do formy wyjściowej. W materiale źródłowym zarejestrowano także inne odpowiedniki, m.in. polskie *w dodatku...* (5,3%) oraz czeskie *a navíc...* (8,5%).

Oba te warianty mogą pełnić funkcję synonimiczną wobec jednostki podstawowej, na co wskazuje ich nieznacznie niższy odsetek udziału w ogólnej liczbie poświadczeń w porównaniu z jednostkami o najwyższej frekwencji użycia. Na podstawie przeprowadzonej analizy można stwierdzić, iż za triadę przekładową należy uznać zestawienie *ко всему прочему... – i do tego (wszystkiego)... – ještě ke všemu...*

W kolejnym przypadku analizie poddano jednostkę *да еуѣ*, która w materiale polskim została najczęściej oddana jako *I do tego...* (3,2%), natomiast w przekładach na język czeski jako *A navíc...* (11,3%).

Jak wskazano we wcześniejszych analizach, oba te translaty zostały uznane za adekwatne ekwiwalenty innych jednostek, co nie wyklucza ich stosowania wymiennie jako synonimów.

Uwzględniając jednak przyjęte założenie poszukiwania form maksymalnie zbliżonych pod względem strukturalnym i jednocześnie zachowujących to samo znaczenie, za ekwiwalenty pełne należy uznać drugie pod względem frekwencji propozycje odnotowane w poświadczeniach korpusowych, tj. polskie *I jeszcze...* (2,1%) oraz czeskie *A ještě...* (9,7%).

Tabelę nr 3 otwiera czeska jednostka *to není všechno...*, której polskie i rosyjskie ekwiwalenty nie budzą wątpliwości co do ich adekwatności. Wysoki udział formy polskiej *to nie wszystko...* (31,3%) oraz rosyjskiej *это не всё...* (20,4%) w ogólnej liczbie poświadczeń, a także ich bliskie podobieństwo strukturalne, pozwalają uznać je za wzajemne odpowiedniki o wysokim stopniu ekwiwalencji. Analiza konkordancji wykazała ponadto, że zarówno w przekładach polskich, jak i rosyjskich, jednostki te często występowały w wariancie poszerzonym o element *jeszcze / ещё*, co wskazuje na tendencję do wzmacniania funkcji zapowiedzi dalszej treści. W materiale źródłowym odnotowano również ekwiwalenty o niższej frekwencji, takie jak polskie *to nie koniec...* (3%), które przygotowuje odbiorcę na kolejną, zwykle niekorzystną informację, czy rosyjskie *есть ещё кое-что...* (5,1%), zachęcające do wzmożonej uwagi odbiorcy w oczekiwaniu na przekazanie istotnej treści. Na tej podstawie można wyodrębnić trójjęzyczny zestaw ekwiwalentów w postaci:

To není všechno... – To nie wszystko... – Есть ещё кое-что....

Ostatnią z jednostek uwzględnionych w przeprowadzonej analizie jest reprodunkt *A navíc...*, który pojawił się już wcześniej jako element ustanowionej triady:

A w dodatku... – И вдобавок... – A navíc....

Jak pokazują dane z tej tabeli, wskazane formy występują również w obecnym zestawieniu, choć nie zajmują pozycji wśród najczęściej używanych. Wśród jednostek polskich najwyższy odsetek poświadczeń uzyskała forma *Poza tym...* (47,5%), która pomimo funkcji wprowadzenia nowej informacji, ma charakter bardziej neutralny i niekoniecznie wzmacnia treści całej wypowiedzi.

W przekładach na język rosyjski równie wysoki wynik osiągnęło wyrażenie *К тому же...* (47,5%), które, jak można wnioskować z analizy poświadczeń, stanowi najbardziej rozpowszechniony i utrwalony sposób oddawania tego typu konstrukcji w tłumaczeniach.

Przeprowadzona analiza umożliwiła wyodrębnienie następujących triad przekładowych dla wybranych fatyzmów, uznanych za adekwatne pod względem zgodności semantyczno-pragmatycznej:

A w dodatku... – И вдобавок... – A navíc...;

Mało tego... – Мало того... – A nejen to...;

A co więcej... – Более того... – A co víc...;

Ko všemu прочemu... – I do tego (wszystkiego)... – Ještě ke všemu...;

Да ещё – I jeszcze... – A ještě...

To není všechno... – To nie wszystko... – Это не всё....

Zastosowanie metody semantyczno-pragmatycznej po raz kolejny potwierdziło jej przydatność w ustalaniu odpowiedników przekładowych z wykorzystaniem źródeł korpusowych. Choć dane frekwencyjne stanowią istotny element procesu analitycznego, nie mogą być traktowane jako czynnik decydujący, gdyż wysoka częstość użycia nie zawsze znajduje odzwierciedlenie w najwyższym stopniu zgodności formalno-semantycznej.

Warto podkreślić, że w wielu przypadkach w roli wzajemnych odpowiedników pojawiały się jednostki już wcześniej zarejestrowane w analizie, co wskazuje na ich wymiennosc i potencjał synonimiczny. Jednocześnie część form wykazuje szczególnie wysoki stopień zbieżności strukturalnej, co stanowiło kluczowe kryterium w procesie ich doboru. Należy także zaznaczyć, że metoda ta opiera się nie tylko na ustalonej semantyce jednostek, bazie reproduktów wielowyrazowych i źródłach korpusowych, lecz również na wiedzy, intuicji i kompetencjach badacza, a połączenie tych elementów decyduje o jej skuteczności. Podsumowując, zastosowanie metody semantyczno-pragmatycznej ujawnia znaczący potencjał badawczy w zakresie identyfikacji trójjęzycznych ekwiwalentów tłumaczeniowych dla wielowyrazowych fatyzmów i emotywów.

Punkt wyjścia w postaci ustalonej wcześniej bazy jednostek z przypisaną semantyką i funkcją pragmatyczną pozwala na prowadzenie analizy porównawczej, bez konieczności ponownej weryfikacji znaczeń w źródłach zewnętrznych. Metoda umożliwia również ujawnienie formalnych i funkcjonalnych powiązań między jednostkami języka polskiego, rosyjskiego i czeskiego, które wcześniej funkcjonowały w zbiorze jako elementy odrębne, a dopiero w zestawieniu w obrębie tej samej kategorii znaczeniowo-funkcjonalnej zyskują status potencjalnych odpowiedników. Jej zaletą jest również zdolność do wychwycenia wariantów formalnych i synonimicznych, co sprzyja poszerzaniu bazy o nowe jednostki, a także do identyfikacji ekwiwalentów odznaczających się wysoką spójnością strukturalną.

W toku analizy możliwe było wyodrębnienie zarówno par o ścisłej przekładalności, jak i rozwiązań funkcjonalnie adekwatnych, lecz różniących się strukturą lub obrazowością. Istotnym atutem metody pozostaje także możliwość łączenia kryteriów jakościowych i ilościowych, co pozwala nie tylko na ocenę stopnia formalno-semantycznej zgodności, lecz także na weryfikację realnego użycia jednostek w praktyce przekładowej. Jednocześnie należy mieć na uwadze jej ograniczenia, np. wysoka frekwencja w korpusach nie zawsze koreluje z najwyższą odpowiednością funkcjonalną, a sam proces doboru ekwiwalentów wymaga nierzadko odwołania się do własnych kompetencji językowych. Metoda semantyczno-pragmatyczna, stosowana jako uzupełnienie innych procedur, stanowi użyteczne narzędzie w

analizie porównawczej i może wspierać dobór jednostek do słowników przekładowych, łącząc analizę znaczenia i funkcji z uwzględnieniem rzeczywistej praktyki komunikacyjnej.

ROZDZIAŁ VIII

Koncepcja artykułu hasłowego fatyzmów i emotywów

Ostatni rozdział niniejszej pracy to propozycja artykułu hasłowego w słowniku elektronicznym, zaprojektowanego z myślą o wielowyrazowych jednostkach fatycznych i emotywnych. W kontekście konstruowanego słownika należy wspomnieć także o pojęciu makrostruktury, czyli sposobie organizacji haseł. W wersji elektronicznej planowany jest układ alfabetyczny, zapewniający intuicyjne wyszukiwanie jednostek, uzupełniony o rozwiązania typowe dla formatu cyfrowego, takie jak wyszukiwanie pełnotekstowe, filtrowanie według kategorii czy hipertekstowe powiązania między hasłami. Uwagę zwrócono jednak przede wszystkim na mikrostrukturę, kluczową dla opisu jednostek fatycznych i emotywnych oraz w największym stopniu decydującą o funkcjonalności słownika. Szczegółowa koncepcja makrostruktury oraz jej aspekty techniczne zostaną omówione w dalszej części pracy.

Artykuł hasłowy, jako podstawowy element mikrostruktury, stanowi fundament słownika, a jego konstrukcja, przejrzystość i funkcjonalność decydują o wartości i użyteczności całego dzieła.

Jak zauważa Piotr Żmigrodzki, to właśnie wzajemny układ haseł oraz wewnętrzna budowa artykułu hasłowego w największym stopniu wpływają na charakter słownika i przesądzają o jego jakości (Żmigrodzki 2009: 12).

Na znaczenie konstrukcji hasła zwracają uwagę również czescy badacze:

„Jedním z nejdůležitějších činitelů rozhodujících o druhu a typu daného slovníku je právě výstavba hesel. Heslem rozumíme slovo-abstrakci, linguisticky určené, zpracované a charakterisované, a to způsobem, který je nejvhodnější pro druh a typ daného slovníku” (Filipec 1957: 129).

Takie ujęcie podkreśla, że każda mikrostruktura powinna być dostosowana do specyfiki i funkcji danego słownika. Podobne podejście widoczne jest w tradycji rosyjskiej, gdzie artykuł hasłowy jednoznacznie określany jest jako kluczowa jednostka dzieła:

«В лексикографии словарная статья признается основной единицей словаря и представляет собой текст особого рода. Эту особенность можно охарактеризовать как полифункциональность, которая заключается в том, что словарная статья выполняет много различных задач» (Цой 2014: 13).

Opracowanie koncepcji artykułu hasłowego jawi się zatem jako zagadnienie kluczowe zarówno dla tradycyjnych słowników drukowanych, jak i dla słowników elektronicznych, które coraz częściej przejmują główną rolę źródła wiedzy językowej.

Rozwój współczesnej leksykografii wskazuje również na coraz silniejsze zorientowanie na użytkownika, co oznacza, że artykuł hasłowy powinien być konstruowany z uwzględnieniem odbiorcy, jego kompetencji i oczekiwań.

Jak zauważa Sue Atkins, słowniki dwujęzyczne powinny nie tylko wspierać rozumienie tekstów, lecz także ułatwiać ich tłumaczenie i produkcję w języku docelowym, a ponadto dostarczać wiedzy o różnicach między jednostkami w obu językach (Atkins 1996: 526). W tym ujęciu to właśnie użytkownik, a nie sam materiał językowy, wyznacza ramy konstrukcji artykułu hasłowego. W literaturze także podkreśla się, że sama obecność hasła nie wystarcza, aby słownik był w pełni użyteczny.

Tadeusz Piotrowski zauważa, że słowniki nierzadko ze względu na swoją formę „nie są użyteczne nawet przy rozumieniu czytanego tekstu” (Piotrowski 2011a: 50).

Podobne stanowisko zajmuje leksykografia czeska, gdzie akcentuje się konieczność zapewnienia odpowiedniego kontekstu, pozwalającego na właściwe rozpoznanie jednostki leksykalnej. Jak zaznacza się w literaturze:

„Slovníkové výklady identifikují sémantickou platnost lexikální jednotky především z hlediska jazykového systému [...], exemplifikační kontexty ji konkretizují, popř. specifikují z hlediska řečových realizací. Tento dvojčediný úkol hesla [...], je ideálem moderní lexikografie.” (Lexikografický kolektiv ÚJČ 1975: 84).

W praktyce oznacza to, że artykuł hasłowy nie może ograniczać się do wyłącznie zaprezentowania ekwiwalentu, lecz powinien uwzględniać także szersze konteksty użycia jednostki.

Na znaczenie przykładów w strukturze artykułu hasłowego zwracają uwagę zarówno badacze zachodni, jak i językoznawcy rosyjscy.

Manfred Spitzer wykazał, że to właśnie poprzez liczne przykłady możliwym jest przyswajanie zasad języka, często nieuświadomianych, ale stosowanych intuicyjnie (Labocha 2021: 124). W leksykografii oznacza to konieczność uzupełniania definicji i ekwiwalentów bogatym materiałem ilustracyjnym.

W tradycji leksykografii rosyjskiej od dawna podkreślano rolę przykładów: «Для более четкой формулировки значения слова [...] в словарную статью нередко включаются иллюстративные примеры. [...] Так, в Толковом словаре живого велико русского языка В.И. Даля языковые иллюстрации представлены пословицами, поговорками,

прибаутками, загадками, в Толковом словаре русского языка под редакцией Д.Н. Ушакова иллюстрации представлены цитатами из художественных произведений, политической литературы и газет середины XX в. [...]» (Резунова 2008: 49).

Obecność przykładów pozwala więc nie tylko na doprecyzowanie znaczenia jednostki, ale także na ukazanie jej funkcji pragmatycznych. Nie mniej istotne jest jednak zachowanie proporcji pomiędzy obszernością informacji a ich przejrzystością. Zbyt duża liczba danych może sprawić, że artykuł hasłowy stanie się nieczytelny i niepraktyczny w odbiorze. W tym kontekście trafna wydaje się „złota reguła” Mirosława Bańki, który wskazuje:

„Podobnie jak w fizyce czy matematyce za eleganckie uznaje się takie rozwiązanie problemu, które jest jednocześnie skuteczne i proste, tak też w leksykografii elegancja wymaga uzgodnienia sprzecznych dążeń: z jednej strony opis leksykograficzny powinien być wyczerpujący i dokładny (w stopniu adekwatnym do potrzeb i oczekiwań odbiorcy), z drugiej zaś powinien być zwięzły i prosty. (...) Słowniki przeznaczone dla ludzi – w przeciwieństwie do tych, które sporządza się z myślą wyłącznie o zastosowaniach maszynowych – muszą bowiem brać pod uwagę ludzką percepcję informacji, w tym przyjemność estetyczną i intelektualną wynikającą ze spostrzeżenia, że opis czegoś, co wydawało się trudne, jest jednocześnie trafny i zaskakująco prosty” (Bańko 2013: 10–11).

Na tle powyższych rozważań jednostki wielowyrazowe o charakterze fatycznym i emotywnym wymagają odrębnego ujęcia w opisie leksykograficznym. Artykuł hasłowy powinien zawierać podstawowe informacje semantyczne, formalne i pragmatyczne, które umożliwią właściwe użycie konkretnej jednostki w różnych sytuacjach komunikacyjnych, przy jednoczesnym unikaniu nadmiaru szczegółowych danych, utrudniających korzystanie z hasła.

W dalszej części pracy celem będzie zaprojektowanie takiej struktury artykułu hasłowego, która dostarczy użytkownikowi niezbędnych informacji w sposób uporządkowany i funkcjonalny. Wysunięte propozycje będą łączyć własne rozwiązania z doświadczeniami wypracowanymi w dotychczasowych opracowaniach leksykografii jednojęzycznej i wielojęzycznej, stanowiących istotny punkt odniesienia i źródło inspiracji.

8.1 Tendencje w opisie artykułów hasłowych jednostek fatycznych i emotywnych w słownikach jedno- i dwujęzycznych

W podrozdziale omówione zostaną główne tendencje w zakresie konstruowania artykułów hasłowych zarówno w słownikach jednojęzycznych, jak i przekładowych. Choć zasadniczym celem nadal pozostaje zaproponowanie struktury artykułu hasłowego dla słownika przekładowego polsko-rosyjskiego i polsko-czeskiego, nie sposób pominąć wzorców wypracowanych na gruncie leksykografii jednojęzycznej. To właśnie artykuł hasłowy słownika jednojęzycznego stanowi punkt wyjścia i istotny kontekst odniesienia dla praktyki przekładowej, wyznaczając pewne standardy w zakresie organizacji informacji, sposobu definiowania jednostek czy prezentacji kontekstów użycia. Obie praktyki opracowywania artykułów hasłowych w słownikach jednojęzycznych i przekładowych w znacznym stopniu się przenikają, choć różnią się zakresem i funkcją. Słownik jednojęzyczny koncentruje się na szczegółowym opisie jednostek, natomiast słownik przekładowy na ustalaniu ich ekwiwalentów w innym języku. Z tego względu w dalszych rozważaniach podjęta zostanie krytyczna refleksja nad tradycyjnym modelem artykułu hasłowego, przy czym krytyka ta nie ma charakteru wartościującego w sensie negatywnym, lecz raczej diagnostycznym. Celem jest wskazanie tych elementów, które choć stanowią istotny fundament dotychczasowej praktyki, wymagają modyfikacji, rozwinięcia lub nowego ujęcia w kontekście opisu jednostek wielowyrazowych, zwłaszcza fatycznych i emotywnych. Jako reprezentatywny materiał analityczny posłużą słowniki omówione w rozdziale szóstym (zob. Rozdział 6. Leksykografia przekładowa), uwzględniające zarówno jednojęzyczne opracowania języka polskiego, rosyjskiego i czeskiego, jak i słowniki przekładowe polsko-rosyjskie oraz polsko-czeskie. Analiza ich mikrostruktury pozwoli uchwycić wspólne cechy tradycyjnych rozwiązań oraz wskazać obszary wymagające dostosowania do potrzeb nowoczesnej leksykografii wielojęzycznej.

W zdecydowanej większości analizowanych opracowań leksykograficznych dominuje tradycyjny **papierowy format słownika**. Było to rozwiązanie w pełni uzasadnione w okresie, gdy technologie cyfrowe dopiero zaczynały wkraczać do praktyki leksykograficznej. Obecnie jednak wyłączny format papierowy istotnie ogranicza potencjał nowych słowników. Nośnik papierowy nie pozwala na dynamiczną aktualizację danych, wprowadzenie rozbudowanych kontekstów pragmatycznych, wielowarstwowych przykładów użycia czy elementów multimedialnych, które w formacie elektronicznym mogą znacząco zwiększać funkcjonalność artykułu hasłowego. Ograniczenia te szczególnie widoczne są w przypadku słowników rejestrujących jednostki wielowyrazowe, gdzie precyzyjny opis wymaga odwołania do

bogatego materiału ilustracyjnego oraz wskazania typowych kontekstów komunikacyjnych. Dobrym przykładem jest seria opolskich idiomatykonów, w której nowatorska koncepcja opisu jednostek nie została w pełni wykorzystana, ponieważ hasła sprowadzają się do prostego zestawienia polskiej jednostki z jej rosyjskim odpowiednikiem, bez komentarzy pragmatycznych, przykładów użycia czy szczegółowych objaśnień.

Podobnie w *Nowym słowniku rosyjsko-polskim / polsko-rosyjskim* Jana Wawrzyńczyka, mimo obecności licznych kolokacji i jednostek wielowyrazowych, brak kontekstów użycia ogranicza jego praktyczną przydatność, zwłaszcza w sytuacjach wymagających interpretacji znaczenia w określonym rejestrze komunikacyjnym.

Jeszcze wyraźniej problem ten widoczny jest w czeskim opracowaniu *Slovník české frazeologie a idiomatiky* Františka Čermáka, który mimo niezwykle rozbudowanej struktury artykułu hasłowego i starannie opracowanych opisów, nie może w pełni oddać aspektów prozodycznych czy intonacyjnych badanych jednostek. W przypadku użytkownika uczącego się języka możliwość odsłuchania nagrania, zobaczenia nagrania wideo czy skorzystania z interaktywnych komentarzy mogłaby znacząco ułatwić przyswajanie kontekstów użycia poszczególnych reproduktów. W takiej perspektywie format elektroniczny jawi się jako nie tylko nowoczesna alternatywa, lecz wręcz konieczność, pozwalająca na elastyczne i wielowymiarowe prezentowanie danych słownikowych.

Problem wielokrotnie już wspomnianego **wyrazocentryzmu** powraca również w kontekście analizy mikrostruktury artykułu hasłowego i stanowi jedno z kluczowych ograniczeń tradycyjnych opracowań leksykograficznych. Na podstawie dokonanych obserwacji można stwierdzić, że dotychczasowa leksykografia rozwija się zasadniczo w trzech nurtach:

- 1) w rejestracji niemal wyłącznie jednostek jednowyrazowych;
- 2) w ujmowaniu jednostek wielowyrazowych jedynie jako uzupełnień haseł głównych, przeważnie w postaci kolokacji, idiomów lub przykładów użycia;
- 3) w opracowywaniu słowników poświęconych w całości jednostkom wielowyrazowym.

Dwie pierwsze tendencje dominują w zdecydowanej większości opracowań i bezpośrednio determinują strukturę artykułu hasłowego. W rezultacie udział jednostek wielowyrazowych w porównaniu z jednowyrazowymi jest znacznie niższy i pozostaje nieproporcjonalny. Taki układ sprawia, że użytkownik może odnieść wrażenie, iż język składa się przede wszystkim z jednostek jednowyrazowych. Tymczasem w rzeczywistej komunikacji to właśnie utrwalone ciągi wielowyrazowe, w tym również reprodukty o funkcji fatycznej i

emotywnej, odgrywają istotną rolę w kształtowaniu relacji i organizacji wypowiedzi. Analiza słowników omówionych w rozdziale pierwszym pokazuje, że nawet jeśli w artykułach hasłowych pojawiają się jednostki wielowyrazowe, są to najczęściej wyłącznie typowe kolokacje lub ustabilizowane idiomy, natomiast reproduktory o charakterze fatycznym i emotywnym są pomijane. Przyczyną pozostaje ich rozmyta struktura, duża wariantowość formalna, przejawiająca się przenikaniem komponentów zmiennych do elementów stałych konstrukcji, oraz wieloznaczność, które utrudniają ich jednoznaczne opisanie w ramach tradycyjnego, linearnego układu hasła. Co więcej, nawet w opracowaniach, które rejestrują takie jednostki, ich liczba jest niewielka. Jednostki wielowyrazowe stanowią tam jedynie około 5–7% całości, a reproduktory istotne z perspektywy niniejszych badań zaledwie 2–3%. Taka tendencja nie tylko ogranicza potencjał informacyjny artykułu hasłowego, lecz także zniekształca obraz rzeczywistego uzusu językowego, w którym utrwalone jednostki wielowyrazowe, w tym fatyzmy i emotywy, odgrywają kluczową rolę w interakcjach komunikacyjnych.

Kolejnym zagadnieniem jest **problem ograniczonej informacji pragmatycznej** w słownikach, które w niewielkim zakresie uwzględniają pojedyncze jednostki wielowyrazowe, w tym fatyzmy i emotywy. Autorzy poszczególnych opracowań skoncentrowali się przede wszystkim na systemowym ujęciu reproduktu, pomijając szerszą charakterystykę sytuacyjną konkretnej jednostki. W efekcie hasła funkcjonują w pewnej izolacji, co znacząco utrudnia zrozumienie ich roli w rzeczywistej komunikacji. Dotyczy to zarówno słowników jednojęzycznych, jak i przekładowych. Przykładem może być uprzednio analizowany *Słownik polszczyzny potocznej*, w którym pod hasłem *ptak* zamieszczono frazę *po ptakach* wraz z krótkim objaśnieniem: ‘koniec, już po wszystkim, skończyło się’ oraz pojedynczym przykładem użycia. Brakuje tu jednak wskazania typowych kontekstów, pozycji składniowej czy jednostek bliskoznacznych, takich jak *no i po zawodach*. Podobną tendencję można zaobserwować w pracy *Толковый словарь русской разговорной речи*, gdzie przy jednostce *легко сказать* ograniczono się do ogólnej następującej informacji: ‘выражает сомнение в возможности того, о чём говорит собеседник’. Co prawda, jest ona uzupełniona kilkoma przykładami użycia, jednak brak w niej pogłębionej analizy funkcji czy wariantów pragmatycznych. Analogiczny problem występuje w pracy *Polsko-český slovník* Karla Olivy, gdzie rozpowszechniony emotyw reakcyjny *a nie mówiłem?* został odnotowany wyłącznie w formie prostego ekwiwalentu *neřikal jsem to?*, bez dodatkowego komentarza. Ograniczony opis tych jednostek jest oczywiście zrozumiały, gdyż analizowane opracowania nie były projektowane z myślą o kompleksowej rejestracji fraz wielowyrazowych. Niemniej jednak, z

uwagi na ich wieloznaczność i istotne funkcje pragmatyczne, bardziej szczegółowe ujęcie wydaje się konieczne, aby użytkownik mógł właściwie zastosować daną jednostkę zarówno w generowaniu tekstu pisanego, jak i w komunikacji ustnej.

Pozostając przy problematyce opisu jednostek wielowyrazowych, warto zwrócić uwagę na słowniki przekładowe, w których rejestrowane są interesujące nas konstrukcje. Istotnym ograniczeniem tychże opracowań jest **brak pogłębionej analizy zestawianych par przekładowych** i ich ekwiwalentów. W żadnym z analizowanych słowników polsko-rosyjskich i polsko-czeskich, w których podawane są odpowiedniki dla jednostek wielowyrazowych, nie pojawia się ich gradacja. Zestawione ekwiwalenty przyporządkowuje się w relacji jeden do jednego, bez wskazania różnic w ich frekwencji, nacechowaniu czy typowych kontekstach użycia. Podawanie wariantów jednostki wyjściowej należy do rzadkości. W nielicznych przypadkach, takich jak słownik *Pragmaticon*, zostało to co prawda uwzględnione, jednak również tam nie towarzyszy temu większa różnorodność odpowiedników. Kwestia wymagająca szczególnego dopracowania dotyczy także kwalifikatorów, zwłaszcza w odniesieniu do jednostek języka mówionego, takich jak fatyzmy i emotywy. Zdecydowana większość tego typu jednostek przynależy do rejestru potocznego lub neutralnego, jednak w ramach samej potoczności wskazane byłoby precyzyjne wyróżnienie jednostek nacechowanych – ironicznych, żartobliwych czy wulgarnych – co umożliwiłoby użytkownikowi pełniejsze zrozumienie ich funkcji oraz zakresu użycia. W niektórych opracowaniach, o których była mowa w rozdziale pierwszym, kwalifikatory wprowadzie się pojawiają, jednak stosowane są wybiórczo i bez konsekwentnej metodologii, co ogranicza ich wartość informacyjną.

Podjęta w niniejszym podrozdziale refleksja nad rozwiązaniami stosowanymi w tradycyjnych artykułach hasłowych pozwala wychwycić zarówno charakterystyczne tendencje, jak i ograniczenia w opisie jednostek wielowyrazowych, w tym fatyzmów i ematywów. Przeprowadzona na podstawie słowników jednojęzycznych oraz przekładowych obserwacja ujawniła kilka powtarzających się zjawisk:

- dominację wyrazocentryzmu i marginalizowanie wielowyrazowych reproductów,
- zbyt ogólnikowy opis jednostek,
- niewystarczającą informację pragmatyczną oraz
- brak pogłębionej analizy ekwiwalentów.

Szczególnie istotnym ograniczeniem okazał się tradycyjny, papierowy format opracowań, który uniemożliwia dynamiczną aktualizację danych, rozbudowanie kontekstów

użycia czy odsyłanie do innych powiązanych znaczeniowo jednostek. Przetawienie się na format elektroniczny otwiera zupełnie nowe możliwości w zakresie organizacji i prezentacji informacji, tj. pozwala na rozbudowywanie opisów, aktualizowanie danych, a także na wykorzystanie materiałów multimedialnych ułatwiających użytkownikowi pełniejsze zrozumienie analizowanych struktur. Pomimo wskazanych ograniczeń omawiane opracowania pozostają cennym punktem odniesienia. Ich struktura i rozwiązania redakcyjne dostarczają wartościowego materiału do prac nad nowoczesnym elektronicznym artykułem hasłowym dostosowanym do specyfiki jednostek fatycznych i emotywnych. W kolejnym podrozdziale uwaga zostanie więc skierowana na te elementy dotychczasowych rozwiązań, które mogą zostać twórczo wykorzystane przy opracowywaniu propozycji artykułu hasłowego spełniającego potrzeby współczesnej leksykografii przekładowej.

8.2 Potencjał dotychczasowych opracowań leksykograficznych w projektowaniu nowoczesnego artykułu hasłowego

Analiza dotychczasowych opracowań leksykograficznych, zarówno jednojęzycznych, jak i dwujęzycznych, ujawnia również szereg rozwiązań w obszarze mikrostruktury, które mogą być przydatne przy projektowaniu nowoczesnego artykułu hasłowego dla wielowyrzowych jednostek fatycznych i emotywnych. Pomimo faktu, iż większość tradycyjnych słowników nie opisuje systematycznie specyfiki reproductów mówionych, ich struktura i układ informacji pokazują dobre praktyki, które można twórczo wykorzystać w leksykografii dwujęzycznej. Potencjał adaptacyjny tychże opracowań wynika z faktu, że oferują sprawdzone sposoby porządkowania treści i prezentacji danych w sposób przejrzysty dla użytkownika. Wyraźne rozdzielanie znaczeń, konsekwentne stosowanie kwalifikatorów, rejestracja kolokacji i przykładów kontekstowych czy system odsyłaczy do jednostek pokrewnych to rozwiązania, które po odpowiednim dostosowaniu, sprawdzą się także w słowniku dwujęzycznym. W projektowanym artykule hasłowym reproductów fatycznych i emotywnych pozwalają one na stworzenie spójnej mikrostruktury, zdolnej pomieścić informacje semantyczne, pragmatyczne i translatoryczne.

Analiza słowników przeprowadzona w rozdziale pierwszym wykazała duże zróżnicowanie jakości i kompletności rozwiązań, co sprawia, że nie wszystkie opracowania w równym stopniu eksponują elementy przydatne do opisu reproductów. Dlatego w niniejszym podrozdziale ujęto jedynie te źródła, które w naszej ocenie, oferują największy potencjał adaptacyjny przy projektowaniu elektronicznego artykułu hasłowego. Są to zarówno słowniki jednojęzyczne (polskie, rosyjskie i czeskie), jak i wybrane słowniki dwujęzyczne, w tym

przekładowe polsko-rosyjskie i polsko-czeskie. Tabela zamieszczona poniżej systematyzuje te źródła i wskazuje konkretne elementy mikrostruktury, które mogą stać się inspiracją dla koncepcji docelowego artykułu. Uwzględnia ona zarówno tradycyjne strategie porządkowania treści, jak i nowoczesne rozwiązania charakterystyczne dla leksykografii elektronicznej, np. hipertekst, integrację nagrań audio i wideo, systemy kolorystyczne czy rozbudowane sieci powiązań semantycznych. Włączenie tych elementów do opisu reproduktów fatycznych i emotywnych umożliwia zatem połączenie pełnego opisu z funkcjonalnością, której oczekują współcześni użytkownicy słowników internetowych. Wyniki tej analizy zostały uporządkowane w tabeli zamieszczonej poniżej, która wskazuje źródła oraz elementy ich mikrostruktury o największym potencjale adaptacyjnym:

Polskie słowniki jednojęzyczne	
Słownik	Rozwiązania o potencjale adaptacyjnym
Stanisław Skorupka. <i>Słownik frazeologiczny języka polskiego</i>	• Systematyczne wyróżnianie kwalifikatorów stylistycznych i pragmatycznych, precyzujących rejestr i zakres użycia jednostki. • Obecność licznych przykładów kontekstowych, często zaczerpniętych z literatury i ilustrujących rzeczywiste funkcjonowanie jednostki w różnych odmianach języka. • Zróżnicowanie typograficzne, uwzględniające m.in. kursywę i nawiasy, ułatwiające lokalizację poszczególnych informacji.
<i>Elektroniczny Wielki Słownik Języka Polskiego</i>	• Wprowadzenie sekcji o składni w artykule hasłowym, np. wskazanie pozycji re produktu w zdaniu. • Krótki komentarz pragmatyczny opisujący rejestr, relacje między rozmówcami lub kontekst komunikacyjny. • System powiązań hipertekstowych do jednostek pokrewnych, wariantów, synonimów i frazeologizmów. • Ilustracje kontekstowe zaczerpnięte z korpusów.
Mieczysław Szymczak <i>Słownik języka polskiego PWN</i>	• Segmentacja i hierarchizacja znaczeń podstawowych, przenośnych oraz wariantów funkcjonalnych poprzez punktację i oznaczenia graficzne.

	Graficzne porządkowanie informacji – spójny system typograficzny (symbole, wyróżnienia, układ) ułatwiający szybkie odnalezienie komentarzy w obrębie artykułu hasłowego.
<i>Inny słownik języka polskiego PWN</i> red. Mirosław Bańko	- Kompleksowy opis znaczeń – staranne ujęcie możliwie pełnego spektrum znaczeń jednostki, zawierające zarówno sensy podstawowe, jak i bardziej szczegółowe lub kontekstowe. - Przystępny język definicji – formułowanie opisów w sposób klarowny, zrozumiały dla użytkownika.

Rosyjskie słowniki jednojęzyczne	
Słownik	Rozwiązania o potencjale adaptacyjnym
<i>Толковый словарь русской разговорной речи</i> ред. Леонид Павлович Крысин	- Uporządkowany wielopoziomowy opis znaczeń – każda jednostka posiada własną definicję wraz z przypisanymi metadanymi (morfologia, składnia, styl, sieć synonimiczną). - Przejrzysta segmentacja i spójna typografia – numeracja znaczeń, wyróżnione bloki (m.in.: DEF, SYN, PHRAS), jednolity układ graficzny. - Wyraźna prezentacja aspektów pragmatycznych – krótkie, trafne charakterystyki funkcji (np. wyrażanie zgody, podziwu, dezaprobaty) oraz profile pragmatyczne poszczególnych znaczeń. - Bogata i zróżnicowana ilustracja kontekstowa – cytaty z korpusów, literatury, filmów i tekstów internetowych z podaniem źródeł i dat.
Александр Ильич Фёдоров <i>Фразеологический словарь русского литературного языка</i>	Precyzyjne wyszczególnienie rejestru i nacechowania – przy jednostkach języka mówionego konsekwentnie wskazuje się rejestr (np.

	<i>разговорное, просторечие</i>) oraz nacechowanie stylistyczne i emocjonalne (np. <i>экспрессивное, ироническое, шутливое</i>).
<i>Большой толковый словарь русского языка</i> ред. Сергей Александрович Кузнецов	• Przejrzysty układ makrostruktury i powtarzalny schemat artykułów hasłowych. • Uwzględnienie wariantów i konstrukcji typowych – np. kolokacje, utrwalone zwroty, użycia przenośne. • Precyzyjne rozróżnienie znaczeń – każde znaczenie jest jasno oddzielone numeracją i wyróżnione w strukturze hasła.
<i>Pragmaticon</i> https://pragmaticon.ruscorpora.ru/	• Wielowymiarowy opis jednostki – przy każdej konstrukcji wskazane są funkcje pragmatyczne, takie jak np. potwierdzenie, zaprzeczenie, upewnienie, wahanie czy dystans • Rozbudowana sieć powiązań – jednostki są powiązane z wariantami konstrukcyjnymi oraz semantycznie zbliżonymi hasłami • Zilustrowanie jednostek kontekstami z mowy potocznej, często w formie dialogów, z dokładnym zapisem struktury • Informacje o typowej intonacji oraz gestach towarzyszących użyciu danej konstrukcji. • Aspekty multimedialne – nagrania audio i wideo powiązane z każdym hasłem. • Kwalifikatory nacechowania emocjonalnego w przypadku niektórych jednostek. • Angielskie odpowiedniki – ekwiwalenty w języku angielskim podawane przy części hasel.

Czeskie słowniki jednojęzyczne	
Słownik	Rozwiązania o potencjale adaptacyjnym
František Trávníček <i>Slovník jazyka českého</i>	• Rozbudowana rejestracja wariantów – jednostki wielowyrazowe odnotowane w obrębie haseł głównych są przedstawiane w wielu swoich wariantach.
<i>Slovník současné češtiny LINGEA</i>	• Zastosowanie kolorów i ikon w ramach do oznaczania wybranych typów informacji • Systematyczna segmentacja znaczeń – podział znaczeń z numeracją i jednolitą strukturą każdego segmentu. • Precyzyjne kwalifikatory – konsekwentne oznaczanie rejestru, dziedziny użycia lub nacechowania stylistycznego za pomocą skrótów (np. <i>expr.</i>, <i>hud.</i>, <i>gastr.</i>).
<i>Akademický slovník současné češtiny</i>	• Informacja o wymowie – zapis fonetyczny jednostki wraz z nagraniem audio • Opis fleksji – podanie form odmiennych w różnych przypadkach • Kwalifikatory stylistyczne – oznaczenie rejestru i nacechowania, np. <i>kolokwialní</i>, <i>expresivní</i>. • Synonimia – lista synonimów i jednostek pokrewnych z aktywnymi odsyłaczami. • Ilustracja kontekstowa – przykłady użycia jednostki w naturalnych zdaniach.
František Čermák <i>Slovník české frazeologie a idiomatiky</i>	• Opis intonacji jednostki – charakterystyka typowych wzorców prozodycznych z uwzględnieniem schematów charakterystycznych dla różnych typów wypowiedzi, np. <i>věta oznamovací</i>, <i>otázka zjišťovací</i> • Charakterystyka stylistyczna – oznaczenie rejestru i nacechowania jednostki, np. <i>kolokviální</i>, <i>expresivní</i>, <i>neutrální</i>. • Informacja o kontekście użycia – wskazanie sytuacji komunikacyjnych

	i typowych konfiguracji, w których jednostka występuje. • Synonimia i ekwiwalenty wielojęzyczne – prezentacja jednostek bliskoznacznych oraz, w przypadku wybranych haseł, ekwiwalentów w pięciu językach obcych.

Dwujęzyczne słowniki polsko-rosyjskie	
Słownik	Rozwiązania o potencjale adaptacyjnym
Dymitr Hessen i Ryszard Stypuła <i>Большой польско-русский словарь</i>	• Rozbudowana segmentacja znaczeń – wyraźny podział na znaczenia główne i poboczne, z numeracją i konsekwentnym układem mikrostruktury.
Elena Nevzorova-Kmech <i>Polsko-rosyjski słownik frazeologiczny gwary młodzieżowej</i>	• Rozbudowany system kwalifikatorów – konsekwentne oznaczanie rejestru, nacechowania i kontekstu użycia jednostki, np. żart., wulg., eufem., pot., lekcew., ucz. stud. • Oznaczenie funkcji pragmatycznej – w niektórych jednostkach wskazana jest ich funkcja komunikacyjna, np. przywitanie.
<i>Podręczny idiomatykon polsko-rosyjski</i> (2006–2019) red. Wojciech Chlebda	• Rozbudowany system symboli graficznych – konsekwentne stosowanie zróżnicowanych symboli graficznych w obrębie artykułów hasłowych, np. wskazanie komponentów fakultatywnych przy użyciu nawiasów łamanych „<...>”. • Oznaczenie akcentu wyrazowego poprzez stosowanie kreski pod sylabą akcentowaną. • Selektywne oznaczanie nacechowania stylistycznego tam, gdzie ma ono kluczowe znaczenie dla interpretacji jednostki.

Dwujęzyczne słowniki polsko-czeskie	
Słownik	Rozwiązania o potencjale adaptacyjnym
Teresa Zofia Orłoś <i>Wielki czesko-polski słownik frazeologiczny</i>	- Dwustronna prezentacja ekwiwalentów – każda jednostka czeska ma wskazany polski odpowiednik, a każda jednostka polska – czeski. - Konsekwentne stosowanie kwalifikatorów – przy jednostkach pojawiają się oznaczenia rejestru i nacechowania stylistycznego, np. <i>pot.</i>, <i>kniž.</i>, <i>vulg.</i>. - Przykłady kontekstowe – jednostki ilustrowane są zdaniami w języku czeskim z równoległymi tłumaczeniami na język polski. - Obecność polsko-czeskiego indeksu umożliwiającego wyszukiwanie konkretnych jednostek. - Rejestracja jednostek wielowyrazowych związanych ze współczesnymi realiami i językiem potocznym, w tym elementów slangu młodzieżowego.
Karel Oliva <i>Polsko-český slovník</i>	- Wyszczególnienie kwalifikatorów – konsekwentnie stosowane oznaczenia stylistyczne (np. <i>zdr.</i>, <i>przen.</i>, <i>zast.</i>, <i>hovor.</i>). - Systematyczna segmentacja znaczeń – znaczenia jednostek leksykalnych są numerowane, a poszczególne użycia wyraźnie rozdzielone. - Obecność wariantów i form pokrewnych – dla części haseł notowane są warianty formalne i znaczeniowe.
Janusz Siatkowski, Mieczysław Basaj <i>Słownik czesko-polski</i>	- Jasna i konsekwentna organizacja znaczeń – poszczególne znaczenia wyodrębnione są poprzez numerację oraz zróżnicowane znaki interpunkcyjne i typograficzne. - Rozbudowany system kwalifikatorów stylistycznych i dziedzinowych – przy jednostkach pojawiają się oznaczenia rejestru i nacechowania stylistycznego.

	Wykorzystanie elementów graficznych – m.in. rombów, kursywy i znaków interpunkcyjnych, które zwiększają przejrzystość artykułu hasłowego.
<i>Internetowy słownik Lingea</i> https://slovniki.lingea.pl/	- Kolorystyczne wyróżnienie elementów hasła – wyraz hasłowy, część mowy, znaczenia i przykłady są odróżnione konkretnymi kolorami. - Możliwość przechodzenia między hasłami pokrewnymi, np. kliknięcie kolokacji lub wyrażenia przenosi do innego artykułu hasłowego - Przy jednostkach systematycznie podawane są typowe połączenia wyrazowe i przykłady ich użycia. - Integracja różnych rejestrów – w jednym artykule zestawiane są ekwiwalenty neutralne, potoczne, ekspresywne, a także wulgarne.

Obserwacja zestawionych w tabelach elementów mikrostruktury ujawnia szereg rozwiązań, które mogą zostać wykorzystane przy projektowaniu nowoczesnego artykułu hasłowego dla wielowyrazowych jednostek fatycznych i emotywnych. Na pierwszy plan wysuwają się te komponenty, które porządkują treść artykułu hasłowego i zapewniają spójny, przejrzysty układ informacji. W wielu słownikach, zarówno jednojęzycznych, jak i dwujęzycznych, konsekwentnie stosowana jest **segmentacja znaczeń**. Widoczna jest ona w pracach takich jak *Słownik języka polskiego PWN* pod redakcją Szymczaka, *Большой толковый словарь русского языка* pod redakcją Kuzniecowa oraz *Polsko-český slovník* Karla Olivy, gdzie znaczenia jednostki są numerowane i hierarchizowane. Takie uporządkowanie nie tylko sprzyja klarowności opisu, lecz także umożliwia wprowadzanie dodatkowych warstw informacji, np. wariantów znaczeniowych czy niuansów pragmatycznych, bez utraty czytelności struktury. W niektórych słownikach, jak np. w *Słowniku czesko-polskim* Siatkowskiego i Basaja, zastosowano dodatkowe **symbole graficzne**, które porządkują mikrostrukturę i pozwalają użytkownikowi szybciej zlokalizować potrzebne dane. Istotnym elementem, obecnym w wielu opracowaniach, jest również **system kwalifikatorów stylistycznych i pragmatycznych**, precyzujących rejestr i zakres użycia jednostki. Konsekwentne oznaczanie nacechowania stylistycznego oraz rejestru obserwuje się w opracowaniach takich jak *Słownik frazeologiczny języka polskiego* Skorupki,

Фразеологический словарь русского литературного языка Fiodorowa oraz *Wielki czesko-polski słownik frazeologiczny* Teresy Orłoś. W pracy *Akademický slovník současné češtiny* i w internetowym słowniku *Lingea* kwalifikatory zostały dodatkowo uzupełnione skrótami dziedzinowymi lub neutralnymi oznaczeniami stylistycznymi, co podnosi precyzję opisu i ułatwia użytkownikowi wybór właściwego odpowiednika w zależności od kontekstu komunikacyjnego. Tego typu rozwiązania stanowią istotną inspirację dla projektowania artykułu hasłowego reproductów, w których odróżnienie rejestrów, np. ekspresywnego czy wulgarneho, ma kluczowe znaczenie dla poprawnej interpretacji i użycia jednostki.

Kolejnym elementem często pojawiającym się w analizowanych słownikach jest **rejestracja wariantów formalnych i znaczeniowych**. Rozwiązanie to uwzględnia dzieła takie jak *Polsko-český slovník* Karla Olivy oraz *Slovník jazyka českého* Trávníčka, w których obok form podstawowych prezentowane są formy wariantywne, pozwalające uchwycić pełniejsze spektrum funkcjonowania jednostki w realnym uzusie. W słownikach dwujęzycznych, takich jak seria *Idiomatikonów polsko-rosyjskich*, zarejestrowano ponadto warianty komponentów jednostek wielowyrazowych, co zwiększa praktyczną wartość artykułu hasłowego dla użytkownika. Na uwagę zasługuje również konsekwentne **umieszczanie przykładów kontekstowych**, które ilustrują rzeczywiste funkcjonowanie jednostki. W wielu opracowaniach przykłady te są zróżnicowane stylistycznie i pragmatycznie, co można zaobserwować w publikacjach takich jak *Толковый словарь русской разговорной речи* Krysina czy *Wielki elektroniczny słownik języka polskiego*. Z kolei w *Wielkim czesko-polskim słowniku frazeologicznym* Orłoś jednostki frazeologiczne zilustrowano równoległymi przykładami w obu językach, co wzmacnia nie tylko funkcjonalność hasła, ale także sprzyja precyzyjnej identyfikacji ekwiwalentów w kontekście dwujęzycznym. Nowoczesne opracowania leksykograficzne pokazują także potencjał **rozwiązań multimedialnych i hipertekstowych**. Szczególnie wyraźnie widać to w projekcie *Pragmaticon*, w którym artykuły hasłowe są wzbogacone o nagrania audio i wideo, a także informacje o typowej intonacji i gestach towarzyszących użyciu jednostki. Hipertekstowa sieć powiązań pozwala użytkownikowi płynnie przechodzić między wariantami konstrukcyjnymi czy semantycznie pokrewnymi jednostkami, co znacząco zwiększa funkcjonalność słownika. Podobne rozwiązania, choć w nieco prostszej formie, zastosowano również w internetowym słowniku *Lingea*, gdzie kliknięcie wyrażenia przenosi do artykułu hasłowego innej jednostki. W części słowników pojawiają się także elementy **wielojęzyczności**. *Pragmaticon* i *Slovník české frazeologie a idiomatiky* Čermáka są zasadniczo opracowaniami jednojęzycznymi, jednak w ich strukturze

wybiórczo występują komponenty wielojęzyczne. *Pragmaticon* zawiera wyłącznie odpowiedniki angielskie, a w słowniku Čermáka ekwiwalenty w pięciu językach obcych.

W tym miejscu warto wskazać na elementy rzadziej spotykane, ale szczególnie interesujące z perspektywy projektowanego słownika reproduktów fatycznych i emotywnych. Zarówno *Slovník české frazeologie a idiomatiky* Čermáka, jak i *Pragmaticon* rejestrują informacje o **intonacji jednostek**, co ma znaczenie przy opisie reproduktów wymagających wskazania charakterystycznego wzorca prozodycznego. Dodatkowo w tym drugim zamieszczono **opisy gestów** towarzyszących użyciu niektórych konstrukcji, co może stanowić inspirację do wzbogacenia przyszłego artykułu hasłowego o komponent pozajęzykowy.

Podsumowując, analiza zestawionych rozwiązań pokazuje, że projektowanie nowoczesnego artykułu hasłowego dla reproduktów fatycznych i emotywnych wymaga uwzględnienia kilku kluczowych elementów:

- przejrzystej segmentacji znaczeń i logicznej hierarchii informacji,
- konsekwentnego systemu kwalifikatorów,
- rejestracji wariantów formalnych i znaczeniowych,
- bogatej ilustracji kontekstowej, a także
- rozwiązań multimedialnych i hipertekstowych.

Zaprojektowana w taki sposób mikrostruktura nie tylko odzwierciedla rzeczywiste funkcjonowanie jednostek w języku, lecz także zwiększa funkcjonalność słownika, czyniąc go narzędziem przydatnym zarówno dla praktyki przekładowej, jak i badań językoznawczych. Uwzględniając zarówno rozwiązania zaczerpnięte z dotychczasowych opracowań leksykograficznych, jak i autorskie propozycje wynikające z analizy zgromadzonego materiału, w kolejnym podrozdziale zaprezentowana zostanie koncepcja projektu słownika, rejestrującego jednostki fatyczne i ematywne.

8.3 Koncepcja makro- i mikrostruktury słownika jednostek fatycznych i emotywnych

W niniejszym podrozdziale podjęta zostanie próba opracowania koncepcji słownika jednostek fatycznych i emotywnych w wersji elektronicznej. Ujęcie to należy traktować jako prototypowe, mające charakter projektowy i stanowiące punkt wyjścia do dalszych prac nad strukturą i funkcjonalnością słownika. W zamyśle autora słownik ma rejestrować wielowyrazowe jednostki fatyczne i ematywne języka mówionego na podstawie autentycznej bazy materiałowej, uwzględniającej dane z korpusów, internetu oraz nagrań audio i audiowizualnych. Od początku projektowany jako narzędzie elektroniczne słownik miałby

oferować możliwość stałej aktualizacji i rozbudowy zasobu, co pozwoliłoby na bieżące uzupełnianie bazy przykładów i wariantów jednostek. Koncepcja ta łączy wnioski z analizy dotychczasowych opracowań leksykograficznych z rozwiązaniami autorskimi, tworząc model możliwy do dalszego rozwijania i dostosowywania do specyfiki jednostek fatycznych i emotywnych.

MAKROSTRUKTURA

Makrostruktura, rozumiana jako ogólny układ i organizacja haseł w słowniku, stanowi nadrzędny poziom konstrukcyjny opracowania leksykograficznego, determinujący sposób dostępu do zgromadzonego materiału. W projektowanym słowniku jednostek fatycznych i emotywnych ma ona charakter tradycyjny, podporządkowany potrzebie przejrzystości i intuicyjnego porządkowania treści. Kluczowym założeniem jest, aby rozwiązania przyjęte w obrębie makrostruktury były przystępne i czytelne dla użytkownika, ułatwiając szybkie dotarcie do poszukiwanych jednostek. Poniżej przedstawiono wstępny zarys koncepcji makrostruktury, uwzględniający główne elementy planowanego układu:

Słownik fatyzmów i emotywów

1



2

3

A nie mówiłem?

Aż dziw bierze

Boże Kochany

Całe szczęście

Daj spokój I

5

Jaja sobie robisz?

Koniec końców

Krótko mówiąc

Mało tego

Niech to szlag!

Prawdę mówiąc

Sam nie wiem

4

Wpisz zapytanie wielowyrzowe

6



*Wyszukiwanie
szczegółowe*

7

1. **Tytuł i opis słownika** – element ten pełni zarówno funkcję identyfikacyjną, jak i informacyjną. Po jego aktywowaniu użytkownik uzyskuje dostęp do krótkiego opisu wyjaśniającego, czym są jednostki fatyczne i emotywyne. W opisie tym w przystępny sposób przedstawia się, że fatyzmy to formuły służące otwieraniu, podtrzymywaniu lub zakończeniu kontaktu, np. *moim zdaniem; chciałoby się powiedzieć, że...; poczekaj, poczekaj...*, natomiast emotywy to wyrażenia reakcyjne sygnalizujące emocje, takie jak oburzenie, obojętność, zaskoczenie, niedowierzanie czy radość, np. *chyba sobie żartujesz; nawet o tym nie myśl!; co za ulga!*. Kluczowe jest, aby opis ten został sformułowany w sposób zrozumiały dla przeciętnego użytkownika, co umożliwi mu szybkie zorientowanie się, do jakiego typu formuł odnosi się zawartość słownika.

2. **Wybór kierunku wyszukiwania (PL-RUS/PL-CZ)** – funkcja ta umożliwia ukierunkowanie zapytania na wybrany zasób dwujęzyczny. Po aktywowaniu odpowiedniego symbolu języka słownik przeszukuje bazę danych w obrębie wskazanej pary językowej, udostępniając ekwiwalenty w danym języku dla wprowadzonej jednostki.

3. **Sugestie wyszukiwania** – element ten pełni funkcję dynamicznego podpowiadania jednostek podczas wpisywania zapytania. Po wprowadzeniu początkowego ciągu znaków system analizuje zasób słownika i w czasie rzeczywistym generuje listę propozycji jednostek zaczynających się od podanego fragmentu. Rozwiązanie to usprawnia proces wyszukiwania oraz umożliwia szybszy dostęp do właściwych reproduktów.

4. **Lista alfabetyczna jednostek** – element ten pełni funkcję przeglądu wszystkich zarejestrowanych jednostek fatycznych i emotywnych, ułożonych w porządku *a fronte*. Użytkownik ma możliwość przewijania listy oraz bezpośredniej eksploracji zasobu słownika bez konieczności formułowania precyzyjnych zapytań.

5. **Indeks wieloznaczności jednostki** – Element ten służy identyfikacji różnych znaczeń danego reproduktu. W przypadku jednostek, które występują w różnych kontekstach komunikacyjnych i realizują odmienne funkcje pragmatyczne, każdemu z ich znaczeń przypisywany jest odrębny numer w porządku rzymskim (I, II, III itd.). Takie rozwiązanie umożliwia precyzyjne rozróżnienie znaczeń semantycznych i pragmatycznych oraz ułatwia

użytkownikowi dostęp do odpowiedniego kontekstu użycia jednostki, a także do adekwatnego ekwiwalentu w językach docelowych.

6. Pole wyszukiwania jednostki – element ten umożliwia bezpośrednie wprowadzenie pełnej formy re produktu, którą użytkownik chce odnaleźć. W przypadku, gdy wpisana jednostka znajduje się w zasobie słownika, system automatycznie przekierowuje użytkownika do odpowiadającego jej artykułu hasłowego. Jeśli natomiast dana forma nie jest zarejestrowana, mechanizm analizy zapytania identyfikuje jednostki pokrewne lub zawierające kluczowe elementy wpisanego wyrażenia, umożliwiając odnalezienie najbardziej zbliżonych odpowiedników. Dzięki temu rozwiązaniu proces wyszukiwania staje się bardziej elastyczny i dostosowany do różnorodnych sposobów formułowania zapytań przez użytkowników.

7. Wyszukiwanie szczegółowe – element ten umożliwia zaawansowaną filtrację zasobu słownika na podstawie wybranych kryteriów opisujących jednostki fatyczne i emotywnie. Użytkownik może określić m.in. typ jednostki (fatyzm lub emotywny), kategorie znaczeniowe, długość konstrukcji, poziom nacechowania stylistycznego czy inne właściwości, które zostaną zdefiniowane w mikrostrukturze artykułu hasłowego. Po zadaniu odpowiednich parametrów system generuje listę jednostek spełniających wyznaczone kryteria, co pozwala na bardziej pogłębioną eksplorację zasobu, szczególnie przydatną w badaniach językoznawczych.

Przedstawiona koncepcja makrostruktury stanowi wstępny model organizacji słownika jednostek fatycznych i emotywnych w wersji elektronicznej. Uwzględnia ona zarówno tradycyjne rozwiązania leksykograficzne, takie jak układ *a fronte* i przejrzystą listę haseł, jak i mechanizmy charakterystyczne dla formatu cyfrowego, np. dynamiczne podpowiedzi, wyszukiwanie szczegółowe czy ukierunkowanie zapytań na wybrane pary językowe. Celem tak zaprojektowanej makrostruktury jest połączenie intuicyjnej obsługi dla przeciętnego użytkownika z funkcjonalnością odpowiadającą potrzebom bardziej wymagających odbiorców, w tym językoznawców. Przedstawiony model stanowi zatem punkt wyjścia do prezentacji schematu mikrostruktury artykułu hasłowego, która w największym stopniu decyduje o wartości informacyjnej i praktycznej słownika.

MIKROSTRUKTURA

Mikrostruktura, rozumiana jako wewnętrzny układ i sposób prezentacji informacji w obrębie artykułu hasłowego, stanowi zasadniczy element funkcjonalności słownika. W projektowanym słowniku reproduktów fatycznych i emotywnych mikrostruktura została zaprojektowana jako rozwiązanie z jednej strony dostępne i zrozumiałe dla przeciętnego użytkownika, z drugiej – wzbogacone o informacje o charakterze bardziej specjalistycznym, istotne dla badań językoznawczych czy praktyki przekładowej. Przedstawione poniżej propozycje należy traktować jako szkice, które mogą zostać wykorzystane i rozwinięte w przyszłych pracach nad słownikiem reproduktów. Rozwiązania te mają na celu połączenie przejrzystości i intuicyjności z możliwością pogłębionego opisu jednostek, uwzględniającego m.in. dane o funkcjach pragmatycznych, kwalifikatorach, wariantach, przykładach użycia, a także elementach multimedialnych. Poniżej zaprezentowano dwa modele mikrostruktury: pierwszy odnoszący się do jednostki fatycznej, drugi zaś - do jednostki ematywnej. Każdy z nich został opatrzony analogicznym opisem numerycznym poszczególnych elementów, umożliwiającym systematyczne omówienie struktury i wskazanie jej potencjalnych zastosowań:

2
Mówiąc ogólnie / Tak ogólnie mówiąc/ Mówiąc ogólniej

1 Ogólnie mówiąc 3 4 5

Otwierające Podtrzymujące Porządkujące Kończące

6
Sygnalizuje uogólnienie lub streszczenie treści wypowiedzi, używany w odniesieniu do większości zagadnień poruszanych w danej sytuacji.

7
A: Jak wam idzie ten projekt?

B: Wszystko przebiega zgodnie z planem, choć mamy drobne opóźnienia w jednym z etapów, ogólnie mówiąc.

8
Synonimy: Mówiąc w skrócie/ Ogólnie rzecz ujmując/ W ogólnym ujęciu

9
Materiał
wideo

Informacje dodatkowe 10

2
Gоворя вообще / Вообще-то говоря/ Но вообще говоря

1 Вообще говоря 3 4 5

Иницирующие Поддерживающие Организующие Завершающие

6
Сигнализирует обобщение или сжатое изложение содержания высказывания, используется в отношении большинства вопросов, затрагиваемых в данной ситуации.

7
A: Как у вас продвигается этот проект?

B: Всё идёт по плану, хотя есть небольшие задержки на одном из этапов, вообще говоря.

8
Синонимы: Говоря в общем/ Говоря в целом

9
Видеоматериал

Дополнительная информация 10

Robisz sobie jaja? / Chyba sobie jaja robisz

I Jaja sobie robisz?

Sprzeciw

Zdenerwowanie

Rozczarowanie

Używane w sytuacjach wyrażania oburzenia, irytacji lub niezadowolienia wobec zachowania lub wypowiedzi rozmówcy. Służy zakwestionowaniu jego postawy bądź słów uznanych za niewłaściwe lub lekceważące.

A: Oddasz mi dziś te pieniądze?

B: Jeszcze nie mam, musisz poczekać do przyszłego tygodnia.

A: Jaja sobie robisz?

Synonimy: *Chyba sobie żartujesz/ Chyba kpisz/ Żarty sobie stroisz?*

Materiał
video

Informacje dodatkowe

Děláš si ze mě srandu? / To si děláš srandu

I Děláš si srandu?

Námítka

Rozčilení

Zklamání

Používá se v situacích vyjadřování rozhořčení, podráždění nebo nespokojenosti vůči chování či výrokům partnera v rozhovoru. Slouží ke zpochybnění jeho postoje nebo slov, která jsou považována za nevhodná či zlehčující.

A: Vrátiš mi dnes ty peníze?

B: Ještě je nemám, musíš počkat do příštího týdne.

A: Děláš si srandu?

Synonyma: *Děláš si legraci?/ Děláš si prdel?*

Videomateriál

Doplňující informace

1. **Wyraz hasłowy** – podstawowa forma jednostki, prezentowana w nagłówku artykułu hasłowego. Spośród wariantów danej konstrukcji wybierana jest forma o najwyższej frekwencji, ustalana na podstawie danych korpusowych, co zapewnia zgodność opisu z rzeczywistym uzusem językowym.

2. **Warianty jednostki** – element mikrostruktury prezentujący alternatywne formy danej jednostki, które zachowują to samo znaczenie i funkcję pragmatyczną, różniąc się jedynie strukturą leksykalną, szykiem wyrazów lub drobnymi modyfikacjami formalnymi. Uwzględnienie wariantów pozwala użytkownikowi rozpoznać i poprawnie interpretować wszystkie powszechnie używane formy danej jednostki.

3. **Kwalifikator** – element mikrostruktury określający stylistyczne, pragmatyczne lub emocjonalne nacechowanie jednostki. W przypadku fatyzmów dominują kwalifikatory neutralne, natomiast przy emotywach częściej pojawiają się oznaczenia rejestru lub wartości emocjonalnej, takie jak *wulg.*, *pot.*, *ekspr.* w języku polskim, *зпыт.*, *нпосм.*, *пазг.* w języku rosyjskim czy *pejor.*, *hovor.*, *iron.* w języku czeskim. Funkcja kwalifikatora polega na ułatwieniu użytkownikowi prawidłowego doboru jednostki do odpowiedniego kontekstu komunikacyjnego.

4. **Odtwarzanie głosowe** – umożliwia odsłuchanie wymowy hasła głównego, co ułatwia opanowanie poprawnego brzmienia i akcentuacji jednostki.

5. **Kategoria** – element mikrostruktury wskazujący przynależność jednostki do jednej lub kilku kategorii funkcjonalnych fatyzmów i emotywów, wprowadzonych i opisanych w rozdziale czwartym. Klasyfikacja ta odzwierciedla funkcję i pozycję jednostki w wypowiedzi (np. otwierającą, podtrzymującą, porządkującą, kończącą). W przypadku niektórych jednostek możliwe jest przypisanie ich do więcej niż jednej kategorii.

6. **Definicja jednostki** – opis znaczenia re produktu podany w formie przystępnej i zrozumiałej, umożliwiający jego poprawne rozpoznanie i użycie.

7. **Przykład użycia** – element ilustrujący funkcjonowanie jednostki w autentycznym kontekście komunikacyjnym. Najczęściej przyjmuje formę dialogu lub krótkiej wypowiedzi,

co pozwala użytkownikowi zrozumieć jej naturalne otoczenie tekstowe, typowe sytuacje użycia oraz możliwe funkcje pragmatyczne.

8. Synonimy jednostki – element mikrostruktury prezentujący jednostki pokrewne semantycznie, które realizują zbliżone funkcje pragmatyczne. W wersji elektronicznej byłyby one wyróżnione i podlinkowane, co umożliwiłoby szybkie przejście do odpowiadających im artykułów hasłowych. Dzięki temu użytkownik może porównać znaczenia oraz zakres użycia jednostek o podobnej wartości komunikacyjnej.

9. Materiał wideo – funkcja umożliwiająca odtworzenie nagrania prezentującego użycie jednostki w autentycznej sytuacji komunikacyjnej, co ułatwia zrozumienie jej znaczenia, sposobu artykulacji, intonacji oraz kontekstu użycia.

10. Informacje dodatkowe – element mikrostruktury udostępniający po aktywacji rozszerzony zestaw danych o jednostce. Zakres prezentowanych informacji miałby charakter zróżnicowany i zależy od specyfiki danego reproduktu, dlatego nie każda jednostka posiadałaby identyczny komplet danych. W tej części mogą się znaleźć m.in. informacje o intonacji, pozycji składniowej w zdaniu, strukturze formalnej (np. liczbie komponentów), gestach towarzyszących, a także dodatkowe przykłady użycia ilustrujące różne konteksty użycia. Rozwiązanie to umożliwi pogłębiony opis jednostki i stanowiłoby szczególnie wartościowe źródło dla językoznawców oraz tłumaczy.

11. Indeks wieloznaczności jednostki – element pełniący tę samą funkcję jak w makrostrukturze, tj. oznaczający kolejne znaczenia jednostek wieloznacznych za pomocą cyfr rzymskich (I, II, III itd.).

Przedstawione w tym rozdziale rozważania, dotyczące koncepcji nowoczesnego artykułu hasłowego dla jednostek wielowyrazowych, łączą tradycyjne rozwiązania leksykograficzne z nowymi możliwościami, jakie oferuje format elektroniczny. Analiza dorobku jedno- i dwujęzycznych opracowań słownikowych pozwoliła wskazać zarówno dobre praktyki, które mogą stanowić inspirację dla przyszłych prac, jak i braki typowe dla klasycznej leksykografii, zwłaszcza w zakresie systematycznego opisu jednostek wielowyrazowych, w tym fatyzmów i emotywów. Wśród szczególnie istotnych rozwiązań wypracowanych w dotychczasowych słownikach należy wymienić m.in.: przejrzystą segmentację znaczeń,

stosowanie kwalifikatorów, rejestrację wariantów oraz bogatą ilustrację kontekstową. Coraz częściej stosowane są również elementy właściwe nowszej praktyce leksykograficznej, takie jak hipertekst, materiały multimedialne czy systemy powiązań semantycznych w opracowaniach elektronicznych. Jednocześnie w wielu pracach wciąż widoczna jest marginalizacja jednostek mówionych, co uzasadnia potrzebę stworzenia słownika ukierunkowanego właśnie na ten typ reproductów wielowyrzowych. Przedstawiona koncepcja słownika jednostek fatycznych i emotywnych ma charakter szkicowy i prototypowy. Zawiera ona propozycję zarówno makro-, jak i mikrostruktury, które zostały zaprojektowane z myślą o połączeniu intuicyjnej obsługi z możliwością pogłębionej analizy jednostek. Makrostruktura, pomimo wzorowania się na tradycyjnym układzie alfabetycznym, przewiduje wykorzystanie mechanizmów charakterystycznych dla słowników elektronicznych, takich jak dynamiczne podpowiedzi, możliwość filtrowania według określonych kryteriów czy wybór kierunku wyszukiwania w ramach par językowych. Z kolei mikrostruktura została rozbudowana o elementy umożliwiające opis nie tylko semantyczny, lecz także pragmatyczny, zawierający hasło główne, definicję, przykłady użycia, kwalifikatory, synonimy, warianty, a także komponenty multimedialne, takie jak nagrania audio i wideo czy opisy gestów i intonacji. Zaprojektowany w taki sposób artykuł hasłowy pozwala uchwycić specyfikę reproductów języka mówionego, które są zjawiskiem trudnym do ujęcia w ramach tradycyjnych metod słownikarskich. Całość rozdziału udowadnia, że opracowanie nowoczesnego słownika jednostek fatycznych i emotywnych jest zadaniem wymagającym nie tylko twórczego wykorzystania istniejących wzorców, ale także wypracowania autorskich rozwiązań dostosowanych do specyfiki tego typu formuł.

Przedstawiona koncepcja nie stanowi projektu zamkniętego, lecz raczej punkt wyjścia do dalszych badań i prac nad leksykograficzną reprezentacją reproductów. Może ona stać się podstawą do rozwijania nowego typu słownika przekładowego, w którym jednostki języka mówionego zostaną opisane w sposób systematyczny i funkcjonalny, z uwzględnieniem ich wieloznaczności, nacechowania stylistycznego i użycia w praktyce komunikacyjnej.

Integracja tradycji leksykograficznej z możliwościami oferowanymi przez technologię cyfrową otwiera drogę do pełniejszego i bardziej użytecznego opisu jednostek fatycznych i emotywnych, zarówno dla przeciętnych użytkowników, jak i specjalistów.

ZAKOŃCZENIE

Niniejsza rozprawa stanowi próbę ujęcia i opisu wielowyrazowych jednostek języka o funkcji fatycznej i ematywnej, zwanych re produktami, które stanowią nieodłączny element codziennej komunikacji. Celem nadrzędnym badań było przedstawienie metod ich wyodrębniania w języku polskim, rosyjskim i czeskim, zaproponowanie sposobów ich przekładu oraz opracowania w formie nowoczesnego słownika przekładowego.

Praca integruje zatem dwa aspekty badań: teoretyczny, koncentrujący się na opisie i klasyfikacji wielowyrazowych jednostek języka oraz praktyczny, uwzględniający wykorzystanie narzędzi korpusowych, metod analitycznych i rozwiązań leksykograficznych.

Przeprowadzone analizy potwierdziły, że jednostki o funkcji fatycznej i ematywnej, mimo iż są powszechnie używane w języku mówionym, przez długi czas pozostawały na marginesie opisu leksykograficznego. W słownikach jednojęzycznych dominowały jednostki jednowyrazowe i nacechowane stylistycznie, natomiast frazy mówione, spontaniczne i sytuacyjnie uwarunkowane pojawiały się jedynie sporadycznie, zwykle jako uzupełnienie haseł głównych. Rejestrowano przede wszystkim te jednostki, które były stosunkowo łatwe do opisanie pod względem formalnym oraz semantycznym.

Współczesna leksykografia dokonuje jednak coraz wyraźniejszego zwrotu ku rejestracji tego typu jednostek, o czym świadczy obecność licznych zwrotów fatycznych i ematywnych w najnowszych opracowaniach zarówno jednojęzycznych, jak i przekładowych. Analiza ich struktury, sposobu definiowania oraz kwalifikatorów pokazuje, że synteza różnych podejść słownikarskich – od tradycyjnych opisów znaczeniowych po nowoczesne ujęcia pragmatyczne – może stanowić skuteczną metodę tworzenia słownika nowego typu: polsko-rosyjskiego i polsko-czeskiego słownika jednostek fatycznych i ematywnych. Słownik taki pozwalałby nie tylko na prezentację ekwiwalentów, lecz także na ukazanie ich funkcji komunikacyjnych.

Potwierdzono hipotezę, zgodnie z którą korpusy językowe stanowią jedno z najskuteczniejszych narzędzi identyfikacji i analizy re produktów. Umożliwiają one obserwację ich rzeczywistego funkcjonowania w tekście, a także pozwalają uchwycić zmienność formalną i semantyczną jednostek. Ich największą zaletą pozostaje szeroki zakres materiału i dostępność zaawansowanych narzędzi wyszukiwawczych, które znacząco usprawniają proces eksploracji danych. Jednocześnie badania wykazały, że korpusy nie są źródłem wystarczającym, zwłaszcza w odniesieniu do jednostek języka mówionego. Choć zawierają obszerne zbiory tekstów, to udział autentycznych zapisów mowy jest w nich ograniczony, a język mówiony nierzadko

reprezentowany jest w postaci stylizowanych dialogów. Dlatego też materiał korpusowy należy traktować jako punkt wyjścia, wymagający uzupełnienia o dane pozakorpusowe – pochodzące ze słowników, tekstów audiowizualnych, filmów czy rejestrów internetowych – które odzwierciedlają aktualne i spontaniczne użycia jednostek. Połączenie tych źródeł pozwoliło na pełniejsze uchwycenie dynamiki współczesnej komunikacji oraz potwierdziło zasadność łączenia analizy ilościowej z analizą jakościową.

Skuteczna identyfikacja i klasyfikacja jednostek fatycznych i emotywnych wymaga zastosowania metod umożliwiających analizę zarówno formalnych, jak i pragmatycznych aspektów wypowiedzi. W badaniach wykorzystano szereg procedur obejmujących m.in.: analizę korpusową, wyszukiwanie przyrostowe, metodę słownikowo-kolokacyjną, a także techniki oparte na pomocnikach formalnych i metaoperatorach.

Propozycje metod wyodrębniania i opisu jednostek wielowyrazowych opracowano na podstawie uprzednio zgromadzonej bazy materiałowej, liczącej około 1500 fatyzmów oraz 5000 emotywów z języka polskiego, rosyjskiego i czeskiego. Zgromadzony materiał stanowił podstawę do testowania opracowanych procedur identyfikacji, klasyfikacji i ustalania ekwiwalentów tłumaczeniowych. Zastosowanie wskazanych sposobów pozwoliło prześledzić rzeczywiste użycie jednostek mówionych w ich autentycznych kontekstach oraz uchwycić wewnętrzne zróżnicowanie znaczeniowe i strukturalne fatyzmów i emotywów. Choć żadna z metod nie umożliwia pełnej automatyzacji procesu, ich połączenie okazało się skuteczne i pozwoliło ograniczyć czasochłonność analizy materiału.

Opracowane w pracy podejście umożliwiło wyodrębnienie głównych typów jednostek wielowyrazowych o funkcji fatycznej i ematywnej w języku polskim, rosyjskim i czeskim. Fatyzmy, zgodnie z przyjętą klasyfikacją, pełnią przede wszystkim funkcję inicjowania kontaktu, jego podtrzymywania oraz zakończenia. Często występują w połączeniu z określeniami adresatywnymi i modalnymi, co pozwala na zasygnalizowanie zmiany tematu rozmowy. Emotywy natomiast koncentrują się na wyrażaniu emocji i postaw nadawcy wobec rzeczywistości, sytuacji komunikacyjnej lub rozmówcy. W analizowanym materiale szczególnie wyraźnie wyróżniły się grupy wyrażające obojętność, zaskoczenie, zdumienie, radość i złość. Warto podkreślić, że granica między jednostkami fatycznymi a ematywnymi ma charakter płynny, a ich funkcje często się przenikają, tworząc obszar przejściowy, w którym struktury o podobnej formie realizują różne intencje komunikacyjne.

Zagadnienie wariantywności stanowiło istotny element opisu wielowyrazowców mówionych. W toku badań potwierdzono, że wiele jednostek ma charakter konstrukcji schematycznych, zawierających elementy stałe oraz miejsca fakultatywne, które mogą być

wypełniane różnymi komponentami bez naruszenia ogólnego sensu frazy. Takie warianty pojawiają się szczególnie często w języku mówionym, w którym spontaniczność wypowiedzi sprzyja modyfikacjom. W opracowaniach leksykograficznych problem ten można rozwiązywać poprzez wskazanie wariantów najbardziej rozpowszechnionych, opartych o dane frekwencyjne, przy jednoczesnym zachowaniu informacji o możliwych modyfikacjach. Dzięki temu artykuł hasłowy może wiernie oddawać dynamikę i zróżnicowanie języka mówionego, nie rezygnując z przejrzystości opisu.

Szczególną trudność w badaniach stanowiło ustalenie ekwiwalentów tłumaczeniowych dla jednostek o wysokim stopniu wieloznaczności. Analiza porównawcza wykazała, że w przypadku fatyzmów i emotywów trudno mówić o pełnej ekwiwalencji przekładowej, ponieważ ich znaczenie i funkcja są ściśle zależne od kontekstu sytuacyjnego. Odpowiedniki w języku docelowym mogą być jedynie najbardziej zbliżone pod względem intencji komunikacyjnej, nacechowania emocjonalnego i potencjału treściowego. Kluczowe znaczenie ma w tym zakresie sytuacyjność, która stanowi podstawę doboru właściwego ekwiwalentu. Identyczny komunikat może w różnych językach przyjmować odmienną formę, jednak w tych samych warunkach komunikacyjnych pełni tę samą funkcję pragmatyczną. Ustalanie odpowiedników wymaga więc analizy funkcji, a nie tylko struktury formalnej z uwzględnieniem ramki pragmatycznej, która określa intencję nadawcy, relację z odbiorcą i ton emocjonalny wypowiedzi.

Jednym z istotnych rezultatów przeprowadzonych analiz jest propozycja modelu artykułu hasłowego w nowoczesnym słowniku przekładowym, który rejestruje jednostki fatyczne i emotywne. Zgromadzony materiał badawczy oraz obserwacje dotyczące funkcjonowania tych jednostek w trzech językach pozwoliły na zaprojektowanie formatu słownika dostosowanego do specyfiki reproduktów. Zaproponowany model ma charakter elektroniczny i umożliwia wielopoziomowe przetwarzanie danych. Struktura artykułu hasłowego uwzględnia komponent formalny, semantyczny, pragmatyczny oraz przekładowy. W obrębie każdego hasła przewidziano miejsce na rejestrację podstawowej formy jednostki, jej wariantów, kwalifikatorów, charakterystyki funkcjonalnej, a także przykładów użycia pochodzących z korpusów i źródeł pozakorpusowych. Integralną częścią wpisu są również ekwiwalenty tłumaczeniowe, opatrzone informacją o ich funkcji i sytuacyjności, co umożliwia użytkownikowi dobór formy adekwatnej do danego kontekstu komunikacyjnego.

Tego typu rozwiązanie jest zgodne ze współczesnymi tendencjami w leksykografii przekładowej, które zakładają integrację danych językowych z informacjami pragmatycznymi. Uwzględnienie potencjału funkcjonalnego i emocjonalnego jednostek sprawia, że słownik

przestaje być jedynie zbiorem ekwiwalentów formalnych, a staje się narzędziem interpretacji komunikacyjnej. Model ten stanowi również propozycję praktycznego zastosowania wyników badań nad językiem mówionym i potwierdza, że nowoczesna leksykografia przekładowa powinna uwzględniać dane pochodzące z różnych źródeł. Każde badanie oparte na materiale językowym wiąże się jednak z pewnymi ograniczeniami. W przedstawionej pracy dotyczyły one przede wszystkim ograniczonej reprezentatywności korpusów i niewielkiego udziału autentycznych zapisów mowy. Uzupełnienie materiału o dane pozakorpusowe pozwoliło zminimalizować dysproporcje między językami i uzyskać bardziej wierny obraz funkcjonowania jednostek mówionych.

Pomimo wspomnianych ograniczeń uzyskane wyniki pozwalają traktować niniejszą rozprawę jako propozycję uporządkowanego modelu analizy jednostek wielowyrazowych o funkcji fatycznej i ematywnej. Opracowane procedury stanowią praktyczne narzędzie służące identyfikacji, klasyfikacji i opisywaniu takich jednostek oraz prezentacji ich w formie nowoczesnego słownika przekładowego. Zaproponowany model może stanowić punkt odniesienia dla dalszych badań nad rejestracją języka mówionego w opracowaniach leksykograficznych, a także przyczynić się do rozwoju metod opisu reproductów w ujęciu wielojęzycznym. Język mówiony, mimo że często marginalizowany, ujawnia bogactwo utrwalonych struktur wielowyrazowych, które odgrywają kluczową rolę w codziennej komunikacji. Analiza materiału z trzech języków słowiańskich pozwoliła na rozpoznanie charakterystycznych typów jednostek fatycznych i ematywnych oraz ukazanie ich złożonej natury semantycznej i pragmatycznej. Rozprawa stanowi tym samym próbę syntetycznego ujęcia problematyki reproductów oraz propozycję nowego ich opisu, wspierającego rozwój współczesnej frazematyki i leksykografii.

BIBLIOGRAFIA

- Atkins S., 1996: *Bilingual dictionaries: Past, Present and Future*. In: *Euralex '96 Proceedings I*. Ed. Martin Gellerstam et al. Göteborg, p. 515–546.
- Bachtin, M. 1986: *Estetyka twórczości słownej*. Red. E. Czaplejewicz. Przeł. D. Ulicka. Warszawa.
- Baker M., 1992: *In other words: a coursebook on translation*. London–New York.
- Bańko M., 2013: *Za mało czy dość? Dylemat leksykografa w dobie rewolucji informacyjnej (na przykładzie słowników frazeologicznych)*. W: *Perspektywy współczesnej frazeologii polskiej. Między teorią a praktyką frazeograficzną*. Red. G. Dziamka-Lenart, J. Liberek. Poznań, s. 9–16.
- Belczyk A., 2007: *Tłumaczenie filmów*. Wilkowice.
- Bell A., Sharma D., Britain D., 2016: *Labov and sociolinguistics: Fifty years of language in social context*. „Journal of Sociolinguistics”, Vol. 20, Issue 42, p. 397–602.
- Białek E., 2009: *Kolokacja w przekładzie. Studium rosyjsko-polskie*. Lublin.
- Bilík J., 2014: *Rozpoznání emocí v textu*. Brno.
- Blagoeva D., Sosnowski W., 2022: *Lakunarność we frazeologii na przykładzie języka bułgarskiego i polskiego*. „Językoznawstwo”, nr 2(17), s. 49–56.
- Bogucki Ł., 2015: *Metodologia badań w przekładzie audiowizualnym*. W: *Metodologie językoznawstwa. Od dialektologii do dialektyki*. Red. P. Stalmaszczyk. Łódź, s. 73–87.
- Bogusławski A., 1976: *O zasadach rejestracji jednostek języka*. „Poradnik Językowy”, s. 356–364.
- Bogusławski A., 1989: *Uwagi o pracy nad frazeologią*. W: *Studia z polskiej leksykografii współczesnej*, T. 3. Red. Z. Saloni. Białystok, s. 13–30.
- Bogusławski A., 1997: *Zagadnienie jednostek przekładowych*. W: *Współczesny język polski i rosyjski. Konfrontacja przekładowa*. Red. A. Bogusławski, J. Mędelska. Warszawa, s. 7–22.
- Borysowski D., 2024: *Opolska idiomatyka w wersji cyfrowej. Wokół słownika Lidé – Média – Komunikace*. „Bohemistyka”, nr 24(3), s. 465–484.
- Brown P., Lai J., Mercer R., 1991: *Aligning Sentences in Parallel Corpora*. „29th Annual Meeting of the Association for Computational Linguistics”, p. 169–176.
- Bühler K., 2004: *Teoria języka*. Przeł. J. Koźbiał. Wstęp K. Polański. Kraków.

Buttler D., 1982: *Znaczenie strukturalne a znaczenie realne stałych związków wyrazowych (Paralele frazeologii i słowotwórstwa)*. W: *Z problemów frazeologii polskiej i słowiańskiej*, T. 1. Red. S. Skorupka. Wrocław, s. 49–56.

Čechová M., 2000: *Čeština - řeč a jazyk*. Praha.

Čermák F., 1992: *Základy lingvistické metodologie*. Helsinki-Juokslahti-Praha.

Čermák F., 2006: *Korpusová lingvistika dnešní doby*. In: *Korpusová lingvistika: Stav a modelové přístupy*. Red. F. Čermák, R. Blatná. Praha, s. 9–18.

Čermák F., 2007: *Frazeologie a idiomatika – česká a obecná*. Praha.

Čermák F., 2010: *Parallel Corpora: The Case of InterCorp*. In: *InterCorp: Exploring a Multilingual Corpus*. Ed. F. Čermák, P. Corness, A. Klégr. Praha, p. 10–17.

Čermák F., Rosen A., 2012: *The case of InterCorp, a multilingual parallel corpus*. „International Journal of Corpus Linguistics”, 17(3), p. 411–427.

Čermák F., 2017: *Korpus a korpusová lingvistika*. Praha.

Červená V. a kol., 1982: *K dějinám vzniku a vydávání Jungmannova Slovníku*. „Naše řeč”, roč. 65, s. 233–247.

Charciarek A., 2010: *Polskie wyrażenia metatekstowe o funkcji fatycznej i ich odpowiedniki czeskie i rosyjskie*. Katowice.

Charciarek A., 2016: *Корпусы текстов как инструмент переводной лексикографии (на примере фразем с компонентом байка и их русских эквивалентов)*. „Slavia Orientalis”, nr 4, T. LXV, s. 743–758.

Charciarek A., 2018: *Využití paralelního korpusu v translatologii (na základě česko-polského InterCorpu)*. „Bohemistika”, nr 19(2), s. 194–216.

Chlebda W., 2010a: *Reprodukty na warsztacie*. W: *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*. Red. W. Chlebda. Opole, s. 7–13.

Chlebda W., 2010b: *Nieautomatyczne drogi dochodzenia do reproduktów wielowyrazowych*. W: *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*. Red. W. Chlebda. Opole, s. 15–35.

Chlebda W., 2011a: *Poszukiwanie odpowiedników przekładowych: punkty wyjścia*. W: *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Red. W. Chlebda. Opole, s. 7–17.

Chlebda W., 2011b: *Ekwiwalencja i ekwiwalenty: między słownikiem a tekstami*. W: *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Red. W. Chlebda. Opole, s. 21–43.

Chlebda W., 2013: *Korpusologia użytkowa dla początkujących i zaawansowanych*. W: *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*. Red. W. Chlebda. Opole, s. 7–15.

Chlebda W., 2018: *Pamięć a język. Zarys Relacji*. W: *Pamięć w ujęciu lingwistycznym. Zagadnienia teoretyczne i metodyczne*. Red. W. Czachur. Warszawa, s. 56–68.

Chlebda W., 2020: *Reprodukowalność, reprodukcja, reprodukt*. W: *Frazeologia a reprodukowalność w teorii i w praktyce komunikacyjnej: problemy - metody analizy – opis*. Red. W. Chlebda, J. Tarsa. Białystok, s. 27–50.

Chlebda W., 2023: *Nieskończona frazeologia. Esej o początkach*. W: *Polsko-rosyjski słownik par przekładowych. Tom zbiorczy Podręcznego idiomatykonu polsko-rosyjskiego (z. 6-10)*. Red. W. Chlebda. Opole, s. 859–898.

Chlebda W., Milutina T.A., *Emotywy*. W: *Idiomatykon 2*. Red. W. Chlebda. Opole, s. 45–54.

Chlumská L., 2014: *Není korpus jako korpus: Korpusy v kontrastivní lingvistice a translologii*. „Časopis pro moderní filologii”, č. 2 (96), s. 221–232.

Czachur W., 2018: *Lingwistyka pamięci. Założenia, zakres badań i metody analizy*. W: *Pamięć w ujęciu lingwistycznym. Zagadnienia teoretyczne i metodyczne*. Red. W. Czachur. Warszawa, s. 7–55.

Czapiga Z., 2015: *O ekspresywności wypowiedzi emotywnych (na materiale języka rosyjskiego i polskiego)*. „Acta Universitatis Lodzensis. Folia Linguistica Rossica”, nr 11, s. 19–27.

Dějiny češtiny. Red. A. A. Fidlerová, R. Dittmann, F. Martínek, K. Voleková, Praha. Dostup z: ucjtk.ff.cuni.cz.

Doroszewski W., 1954: *Z zagadnień leksykografii polskiej*. Warszawa.

Doroszewski W., 1970: *Elementy leksykologii i semiotyki*. Warszawa.

Doroszewski W., 1982: *Język. Myślenie. Działanie: rozważania językoznawcy*. Warszawa.

Drabik-Frączek B., Załazińska A., 2014: *Komunikologia – zarys teoretyczny i problemy praktyczne w świetle kształcenia kompetencji komunikacyjnych*. „LingVaria”, nr 2 (18), t. IX, s. 29–42.

Engels J., 1968: *Les "Gloses de Reichenau" rééditées*. „Neophilologus”, t. 52, nr 4, s. 378–386.

Filipec J., 1957: *Lexikálně sémantická výstavba hesla — ústřední otázka lexikografické práce (Na materiále jednojazyčných slovníků)*. „Slovo a slovesnost”, roč. 18, č. 3, s. 129–150.

Filipec J., Čermák F., 1986: *Česká lexikologie*. Praha.

- Furdal A., 1977: *Językoznawstwo otwarte*. Opole.
- Gajda S., 1987: *Nauka o kulturze języka*. „Polonistyka”, z. 8, s. 581–592.
- Gajda S., 1988: *O pojęciu idiolektostylu*. W: *Język osobniczy jako przedmiot badań lingwistycznych*. Red. J. Brzeziński. Opole, s. 23–34.
- Gale W., Church K., 1993: *A Program for Aligning Sentences in Bilingual Corpora*. „Computational Linguistics”, Vol. 19 (1), p. 75–102.
- Gasek B., 2017: *Polsko-rosyjska leksykografia par przekładowych: stan aktualny i perspektywy*. „Slavia Orientalis”, t. LXVI, nr 1, s. 139–169.
- Goleman D., 1997: *Emocjonalna inteligencja. Proč může být emoční inteligence důležitější než IQ*. Praha.
- Grochowski M., 1982: *Zarys leksykologii i leksykografii. Zagadnienia synchroniczne*. Toruń.
- Grohová M., 2024: *Po stopách tzv. překladatelštiny v audiovizuálním překladu*. Praha.
- Gruszczyński W., 2013: *From multilingual to monolingual dictionaries: A historical overview of Polish lexicography*. „Studies in Polish Linguistics”, 8(4), s. 205–227.
- Grzegorzczak R., 1991: *Problem funkcji języka i tekstu w świetle teorii aktów mowy*. W: *Funkcje języka i wypowiedzi. – Język a Kultura*, T. 4., Red. J. Anusiewicz, R. Grzegorzczak. Wrocław, s. 11–28.
- Grześkowiak R., 2019: *Dawne krakowskie edycje pierwszej księgi przysłów polskich 3. Stanisław Serafin Jagodyński jako redaktor wydania Przypowieści polskich Salomona Rysińskiego z roku 1620*. „Terminus”, t. 21, z. 2 (51), s. 145–195.
- Hanks P., 2013: *Lexicography from Earliest Times to the Present*. In: *The Oxford Handbook of the History of Linguistics*. Ed. K. Allan. Oxford, p. 503–536.
- Hartmann, R. R. K., 1989: *Sociology of the Dictionary User: Hypotheses and Empirical Studies*. In: *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography*, Vol. 1. Ed. F. J. Hausmann, O. Reichmann, H. E. Wiegand, L. Zgusta. Berlin-New York, p. 102–111.
- Haruno M., Ikehara S., Yamazaki T., 1996: *Learning Bilingual Collocations by Word-Level Sorting*. „COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics”, p. 525–530.
- Hebal-Jezińska M., 2013: *Podstawowe zasady korzystania z korpusów przy badaniu języka*. W: *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*. Red. W. Chlebda. Opole, s. 17–30.

Hebal-Jeziarska M., Kaczmarska E., Rosen A., 2016: *Between the devil and the deep blue sea or between users' needs and the compilers' powers: An analysis of the Czech-Polish part of the parallel corpus InterCorp*. W: *Polskojęzyczne korpusy równoległe. Polish-language Parallel Corpora*. Red. E. Gruszczyńska. Warszawa, s. 41–65.

Hirszfeld A., 2015: *Co rządzi obrazem? Powtórzenie w sztukach audiowizualnych*. Kraków

Hladká Z., 2005: *České slovníkářství na cestě k jednojazyčnému výkladovému slovníku*. „Naše řeč”, roč. 88, č. 3, s. 140–159.

Hodura K., 1959: *Z dějin akademického slovníku jazyka českého*. „Naše řeč”, roč. 42, č. 1–2, s. 6–14.

Hoffmannová J., 1996: *Fatická funkce jazyka, konverzace a její žánry*. „Slovo a slovesnost”, roč. 57, č. 3, s. 191–205.

Hrbáček, J., 1983: *Poznámky k funkčním stylům prostě sdělovacímu, hovorovému a konverzačnímu*. „SlavPrag XXVI. Acta Universitatis Carolinae – Philologica 2”, s. 111–114.

Jakobson R., 1960: *Poetyka w świetle językoznawstwa*. „Pamiętnik Literacki LI, 51/2”, przeł. Krystyna Pomorska, s. 431–473.

Jedlička D., 2019: *Úvod do teorie a praxe překladu. Distanční studijní text*. Opava.

Jelínek M., 1996: *Postavení hovorového stylu mezi styly funkčními*. „Slovo a slovesnost”, roč. 27, č. 2, s. 104–118.

Józwiak-Dądela A., 2007: *O tłumaczeniu polskich związków frazeologicznych na język niemiecki na przykładzie wybranych tekstów publicystycznych*. Katowice.

Kaczmarska E., Rosen A., 2014: *Praktyczny przewodnik po korpusie równoległym InterCorp*, W: *Praktyczny przewodnik po korpusach języków słowiańskich*. Red. M. Hebal-Jeziarska. Warszawa, s. 207–231.

Kaczyńska B., 2019: *Wokabularz trydencki jako przykład leksykografii średniowiecznej*. „Studia Źródłoznawcze”, t. LVII, s. 47–97.

Kapusta O., 2008: *Česko-francouzské slovníkářství ve 20. století na pozadí vývoje lexikografie jako vědní disciplíny*. Olomouc.

Kiklewicz A., 2008: *Dwanaście funkcji języka*. „LingVaria”, Rok III, nr 2 (6), s. 9–27.

Klein H. W., 1965: *Lateinisches und Romanisches in den Reichenauer Glossen*. „Nachrichten der Gesellschaft für romanische Sprach- und Literaturwissenschaft”, nr. 34, p. 181–200.

Koller W., 1972: *Grundprobleme der Übersetzungstheorie*. Bern.

Koller W., 1992L *Einführung in die Übersetzungswissenschaft*. 4., völlig neu bearbeitete Auflage. Heidelberg.

Kominicz L., 1971: *Wieloznaczność a funkcja selekcyjno-determinacyjna kontekstu*. Warszawa-Wrocław.

Konečná M., 2017/2018: *Národní parky z pohledu české právní úpravy*. Brno.

Kozera I., 2018: *Семантика и прагматика вторичной имперфективации в современном русском языке на материале корпусного анализа*. Kraków.

Kurdyła T., 2015: *Najważniejsza funkcja „języka”? Uwagi o funkcji fatycznej i fatyczności języka*. „Biuletyn PTJ LXXI”, s. 59–70.

Kuryłowicz B., 2021: *Największe dzieła polskiej leksykografii XVIII wieku (Nowy dykjonarz Michała Abrahama Troca oraz Nowy wielki dykjonarz Piotra Daneta i Dymitra Koli) z perspektywy współczesnych badań historycznojęzykowych: nowe wyzwania i potrzeby badawcze*. „Białostockie Archiwum Językowe”, nr 21, s. 97–107.

Labocha M., 2021: *Glottodydaktyczne aspekty opisu słownikowego kolokacji*. „Roczniki Humanistyczne”, nr 69(10), s. 121–138.

Landau Sidney I., 1989: *Dictionaries: The Art and Craft of Lexicography*. Cambridge-New York.

Lewandowska-Tomaszczyk B., 2005: *Podstawy językoznawstwa korpusowego*. Łódź.

Lewandowska-Tomaszczyk B., 2010: *Introspekcja i intuicja w uczeniu się kategorii językowych*. W: *Interdyscyplinarne studia nad świadomością i przetwarzaniem językowym*. Red. J. Nijakowska. Kraków, s. 41–74.

Lewicki A. M., 1974: *Aparat pojęciowy frazeologii*. W: *Z badań nad literaturą i językiem*. Red. L. Ludorowski i W. Magnuszewski. Warszawa – Poznań, s. 135–151.

Lewicki A. M., 1976: *Wprowadzenie do frazeologii syntaktycznej. Teoria zwrotu frazeologicznego*. Katowice.

Lewicki A. M., 2003: *Studia z teorii frazeologii*. Łask.

Lewicki R., 2022: *Zagadnienia lingwistyki przekładu*. Lublin.

Lexikografický kolektív ÚJČ., 1975: *K výkladovým postupům v jednojazyčné lexikografii*. „Slovo a slovesnost”, 36(2), s. 84–91.

Maćkowiak K., 2020: *Podstawowe problemy teorii świadomości językowej*. „Poradnik Językowy”, nr 2 (771), s. 29–45.

Malinowski B., 1972: *Phatic Communion*. In: *Communication in face-to-face interaction: textbook*. Ed. J. Laver, S. Hutcheson. Harmondsworth, p. 146–152.

Markowski A., 2006: *Kultura języka polskiego. Teoria. Zagadnienia leksykalne*. Warszawa.

Mathesius V., 1982: *Společenské základy krásného hovoru*. In: *Jazyk, kultura a slovesnost*. Red. J. Vachek. Praha, s. 393–396.

Mcenery A., Xiao R., Tono Y., 2006: *Corpus-based Language Studies: An advanced Resource Book*. London–New York.

McEnergy T., Wilson A., 2001: *Corpus Linguistics. An Introduction*. Edinburgh.

Mielczuk I., 1995: *Phrasemes in Language and Phraseology in Linguistics*. In: *Idioms. Structural and Psychological Perspectives*. Red. M. Everaert, E-J. van der Liden, A. Schenk, R. Schreuder. Hillsdale-New Jersey, p. 167–232.

Mukařovský J., 1948: *Kapitoly z české poetiky*. Praha.

Müldner-Nieckowski P., 2005: *Frazeologia poszerzona. Studium Leksykograficzne*. Warszawa.

Narodowy Korpus Języka Polskiego. Red. A. Przepiórkowski, M. Bańko, R. L. Górski, B. Lewandowska-Tomaszczyk. Warszawa 2012.

Nevzorova-Kmech E., 2022: *Перевод субстандартной фразеологии с польского языка на русский (материалы к переводному словарю жаргона на примере романа Дороты Масловской «Польско-русская война под бело-красным флагом»)*. „Językoznawstwo”, nr 1(16), s. 49–60.

Otulak-Komenda M., 2010: *Akty mowy o funkcji fatycznej w wywiadach telewizyjnych*. Poznań.

Parry M., 1930: *Studies in the Epic Technique of Oral Verse-Making*. „Homer and Homeric Style. Harvard Studies in Classical Philology”, t. 41, p. 73–147.

Pęzik P., 2020: *Budowa i zastosowania korpusu monitorującego MoncoPL*. „Forum Lingwistyczne”, nr 7, s. 133–150.

Piegzik W. A., 2020: *Intuicja językowa i jej funkcja oceniająca: na przykładzie języka francuskiego jako obcego*. „Prace Językoznawcze”, t. XXII, nr 1, s. 219–238.

Piegzik W. A., 2022a: *Mechanizmy intuicji językowej*. „Prace Językoznawcze”, nr 1, t. XXIV, s. 135–150.

Piegzik W. A., 2022b: *Intuicja jako zdolność mentalna i metoda badawcza językoznawcy*. „Linguodidactica”, t. XXVI, s. 181–192.

Pierrel J. M., 2006: *Le Trésor de la Langue Française Informatisé : un dictionnaire de référence accessible à tous*. „Revue de l'Université de Moncton”, t. 37, nr 1, p. 3–20.

Piotrowski T., 1994: *Z zagadnień leksykografii*. Warszawa.

Piotrowski T., 2006: *Zrozumieć leksykografię*. Opole.

- Piotrowski T., 2011a: *Ekwiwalencja w słownikach dwujęzycznych*. W: *Na tropach translatorów. W poszukiwaniu odpowiedników przekładowych*. Red. W. Chlebda. Opole, s. 45–69.
- Piotrowski T., 2011b: *Tertium comparationis w przekładoznawstwie*. W: *Metodologie językoznawstwa. Od ontologii do pragmatyki*. Red. P. Stalmaszczyk, Łódź, s. 175–192.
- Poldauf, I. 1962: *Teorie české synonymiky*. „*Naše řeč*”, roč. 45, č. 9–10, s. 305–312.
- Praktyczny przewodnik po korpusach języków słowiańskich*. Pod red. M. Hebał-Jezierskiej. Warszawa 2014.
- Prangiszwili A., 1969: *Psychologia nastawienia*. Warszawa.
- Raskin J., 2000: *The Humane Interface: New Directions for Designing Interactive Systems*. Boston.
- Říčan P., 2009: *Psychologie*. Praha.
- Rosen A., 2016: *InterCorp – a look behind the façade of a parallel corpus*. W: *Polskojęzyczne korpusy równoległe. Polish-language Parallel Corpora*. Red. E. Gruszczyńska, A. Leńko-Szymańska. Warszawa, s. 21–40.
- Sarzyńska J., 2009: *Język intuicji. Rola języka w procesach pamięciowych i intuicyjnych*. „*Rocznik Kognitywistyczny*”, t. 3, s. 133–138.
- Sawicka G., 1995: *Nominacja rzeczownika na przykładzie mowy dzieci przedszkolnych*. Szczecin.
- Sgall P., 2010: *Perspektivy standardní češtiny*. „*Jazykovědné aktuality*”, roč. XLVII, č. 3–4, s. 73–94.
- Sinclair J., 1991: *Corpus, Concordance, Collocation*. Oxford.
- Sinclair J., 2005: *Corpus and Text – Basic Principles*. In: *Developing Linguistic Corpora: a Guide to Good Practice*. Ed. M. Wynne. Oxford, p. 1–15.
- Skorupka S., 1969: *Podstawy klasyfikacji jednostek frazeologicznych*. „*Prace Filologiczne*”, t. 19, s. 219–226.
- Skubalanka T., 1972: *O ekspresywności języka*. „*Annales Universitatis Mariae Curie-Skłodowska. Sectio F: Nauki Filozoficzne i Humanistyczne*”, Vol. XXVII, s. 123–135.
- Stein G., 1981: *The English Dictionary in the 15th Century*. In: *Geschichte der Sprachphilosophie und der Sprachwissenschaft*. Red. J. Trabant. Berlin-Boston, p. 313–322.
- Stěpanova L., 2014: *Frazeologizmy jako nośniki wartościowania a ich międzyjęzykowa ekwiwalencja*. W: *Frazeologia a překlad*. Red. W. Chlebda. Opole, s. 133–139.

- Sujin M. K., 2018: *Lingvistické hedge v češtině a angličtině*. Praha.
- Szczerbowski, T., 2003: *Recenzja Idiomy polsko-rosyjskie: Chlebda W., Gołubiewa A., Wawrzyńczyk J., Wielg T., Wydawnictwo Naukowe PWN, Warszawa, 2003, 272 s.* „Стил”, s. 379–382.
- Szubin E. P., 1974: *Komunikacja językowa a nauczanie języków obcych*. Przekład zbiorowy. Warszawa.
- Szymańska-Galińska K., 2023: *Świadomość językowa a gwara*. „Językoznawstwo”, nr 1(18), s. 171–178.
- Śniatkowski S., 2011: *Świadomość językowa i metajęzykowa a kompetencja lingwodydaktyczna*. „Annales Universitatis Paedagogicae Cracoviensis. Studia Logopaedica”, t. IV, (96), s. 493–504.
- Tambor J., 1991: *O funkcji fatycznej niektórych elementów tekstów mówionych*. W: *Funkcje języka i wypowiedzi. – Język a Kultura*, t. 4., Red. J. Anusiewicz, R. Grzegorzczkowska, s. 177–182.
- Tiedemann J., 2009: *News from OPUS — A collection of multilingual parallel corpora with tools and interfaces*. In: *Recent Advances in Natural Language Processing*. Ed. N. Nicolov, K. Bontcheva, G. Angelova, R. Mitkov. Amsterdam/Philadelphia, s. 237–248.
- Tognini-Bonelli E., 2001: *Corpus Linguistics at Work*. Amsterdam-Philadelphia.
- Tomaszewski T., 1984: *Ślady i wzorce*. Warszawa.
- Urbańczyk S., 1962: *Wokabularz trydencki*. „Język Polski”, t. 42, z. 1, s. 15–29.
- Vachková M., 2019: *Co přináší jednojazyčná lexikografie lexikografii překladové?* „Naše řeč”, roč. 102, č. 1–2, s. 102–118.
- Vajdllová M., 2019: *Proměny slovníkové definice ve slovnících Daniela Adama z Veleslavína (na českém jazykovém materiálu)*. „Studia z Filologii Polskiej i Słowiańskiej”, nr 54, s. 1–22.
- Voleková K., 2013: *Česká lexikografie 15. Století*. Praha.
- Wakulik K., 2023: *O kompetencji interkulturowej tłumaczy filmów animowanych – na przykładzie rosyjskiego i polskiego dubbingu filmów Shrek i Kung Fu Panda*. „Slavia Orientalis”, T. LXXII, nr 3, s. 599–614.
- Walczak B., 1991: *Słownik wileński na tle dziejów polskiej leksykografii*. Poznań.
- Wawrzyniak Z., 1992: *Praktyczne aspekty translacji literackiej na przykładzie języków niemieckiego i angielskiego*. Warszawa.
- Wawrzyniak Z., 2002: *Zasady rozumienia jako warunki konstytutywne tłumaczenia*. W: *Recepcja. Transfer. Przekład*. Red. J. Koźbiał. s. 3946.

Weinreich U., 1969: *Problems in the Analysis of Idioms*. In: *Substance and structure of Language*. Ed. J. Puhvel. Berkeley-Los Angeles, p. 23–81.

Wojtczuk K., 2010: *Wykrzykniki współczesnej polszczyzny w funkcji fatycznej*. „*Conversatoria Linguistica*”, t.4, s. 87–97.

Wołkowski P., 2010: *Biologiczne aspekty teorii zdolności językowej Chomsky'ego*. „*Diametros*” (23), s. 145–161.

Wójcik T., 1983: *Nominacja a problem kontrastów międzyjęzykowych*. W: *Problemy nominacji językowej*, t. 2. Red. M. Blicharski. Katowice, s. 15–22.

Zięba A., 2022: *Параллельный корпус InterCorp в процессе установления польско-русских пар метаоператоров*. „*Przegląd Rusycystyczny*”, nr (1) 177, s. 195–214.

Żmigrodzki P., 2009: *Wprowadzenie do leksykografii polskiej*. Katowice.

Żmigrodzki, P. 2018: 2. *Ogólna charakterystyka słownika. 2.1 Charakter słownika*. W: *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*. Red. P. Żmigrodzki, M. Bańko, B. Batko-Tokarz, J. Bobrowski, A. Czelakowska, M. Grochowski, R. Przybylska, J. Waniakowa, K. Węgrzynek. Kraków, s. 17–24.

Żydek-Bednarczuk U., 2005: *Wprowadzenie do lingwistycznej analizy tekstu*. Kraków.

Абакумова, О., 2020: *Воспроизводимость семантики и структуры пословицы при её употреблении в художественном тексте*. W: *Frazeologia a reprodukowalność w teorii i praktyce komunikacyjnej. Problemy – metody analizy – opis*. Red. W. Chlebda, J. Tarsa. Białystok, s. 11–26.

Азнабаева Л. А., 1989: *Критерии выделения фатических высказываний и особенности их функционирования в художественном тексте*. Москва.

Акимов Г. Н., Вяткина С. В., Казаков В. П. и др., 2013: *Синтаксис современного русского языка: хрестоматия с заданиями*. Санкт-Петербург.

Апресян Ю. Д., 1993: *Лексикографическая концепция НБАРСа*. В: *НБАРС*. Ред. Ю. Д. Апресян, Москва, с. 6–17.

Балли Ш., 2003: *Язык и жизнь*. Москва.

Баранов М. Т., Ладыженская Т. А., Львов М. Р., 1990: *Методика преподавания русского языка*. Москва.

Беляева Л. Н., 2004: *Лексикографический потенциал параллельного корпуса текстов*. В: *Труды международной конференции «Корпусная лингвистика-2004»*. Ред. Л. Н. Беляева. Санкт-Петербург, с. 55–64.

Бо Л., 2015: *Русские экспрессивные синтаксические конструкции как коммуникативные единицы*. Воронеж.

Бородулина Н. Ю., 2007: *Метафорическая репрезентация экономических понятий как объект семиотического анализа*. Тамбов.

Буга А. С., 2011: *Чувство языка: парадигмы исследования*. «Психологическая наука и образование», № 4, с. 1–13.

Быстрова Е.А., Окунева А. П., Шанский Н.М., 1984: *Учебный фразеологический словарь русского языка. Пособие для учащихся Национальных школ*. Ленинград.

Виноградов В., 1977: *Избранные труды. Лексикология и лексикография*. Москва.

Винокур Т. Г., 1993: *Говорящий и слушающий: Варианты речевого поведения*. Москва.

Гаврилов А.С., 2014: *Словарь синонимов и антонимов современного русского языка*. Москва.

Гасек Б., 2020: *Множественные соответствия в «Польско-русском словаре переводных пар»*. «Вестник Волгоградского государственного университета», 19 (2), с. 86–97.

Гоголина Т. В., 2014: *Коммуникемы со значением сомнительности как отражение лингвокреативной деятельности говорящего*. «Психолингвистика в образовании», № 2, с. 181–185.

Горелов И. Н., 1990: *Коммуникация*. В: *Лингвистический энциклопедический словарь*. Ред. В. Н. Ярцев. Москва, с. 233–234.

Городецкий Б. Ю., 1983: *Проблемы и методы современной лексикографии*. В: *Новое в зарубежной лингвистике. Проблемы и методы лексикографии*. Ред. Т. С. Зевахина, Б. Ю. Городецкий. Москва, с. 5–22.

Гребенников А. О., 2017: *Русская лексикография: история и перспективы*. «Философия и гуманитарные науки в информационном обществе», № 1(15), с. 107–120.

Гридина Т. А., 2015: *Языковая интуиция как эвристический вектор детской речи*. «Педагогическое образование в России», № 11, с. 172–179.

Денисов П. Н., 1993: *Лексика русского языка и принципы её описания*. Москва.

Дзенс Н. И., Перевышина И. Р., 2012: *Теория перевода и переводческая практика: с немецкого языка на русский и с русского на немецкий*. Санкт-Петербург.

Добровольский Д. О., Кретов А. А., Шаров С. А., 2005: *Корпус параллельных текстов архитектура и возможности использования*. В: *Национальный корпус русского языка: 2003—2005. Результаты и перспективы*. Москва, с. 263–296.

Довгаль А. В., 2008: *Синтаксические модели эмотивных высказываний в белорусском и русском языках*. В: *Интегративные функции русского языка и культуры в системе*

непрерывного образования Беларуси, России, Украины: материалы Международной научно-практической конференции (16 мая 2008 г.). Ред. Е. Е. Иванов. Могилев, с. 111–115.

Додонов Б. И., 1978: *Эмоции как ценность*. Москва.

Долгова О., 1980: *Синтаксис как наука о построении речи*. Москва.

Дорохова Ю. Э., 2013: *Двуязычная лексикография и перевод*. «Язык как система и деятельность 4. Материалы Международной научной конференции», с. 73–78.

Дубичинский В. В., 2008: *Лексикография русского языка: Учебное пособие*. Москва.

Дымарский М., 2013: *От моделей предложения к моделированию высказывания*. В: *Славянское языкознание. XV Международный съезд славистов*. Ред. А. А. Кибрик. Москва, с. 325–326.

Ейгер Г. В., Безуглая Л. Р., Бабич Е. Н., 2016: *Интуиция как метод научного познания в лингвистике и переводоведении*. «Вісник Харківського національного університету імені В.Н. Каразіна», № 84, с. 158–167.

Жердева М. В., Артюшенко В. М., 2016: *Стемминг и лемматизация в Lucene.net*, «Лесной вестник», № 3, с. 131–135.

Закон судный людям: краткой редакции, ред. акад. М. Н. Тихомиров. Москва 1961.

Захаров В. П., 2005: *Корпусная лингвистика: Учебно-методическое пособие*. Санкт-Петербург.

Иванова Ю., Портнягин Д., 2016: *Бытие языка в тексте*. В: *Универсальное и культурно-специфичное в языках и литературах*. Курган, с. 154–159.

Изард К. Э., 1999: *Психология эмоций*. Санкт-Петербург.

Илинская А. С., 2006: *Знаковая типология языковых единиц, репрезентирующих эмоции в английском языке*. «Ползуновский вестник», № 3, с. 98–105.

История русской лексикографии. ред. Ф. П. Сороколетов. Санкт-Петербург 2001.

Казачихина И. А., 2008: *Коммуникативы как объект описания в учебном пособии словарного типа*. «Известия Российского государственного педагогического университета им. А. И. Герцена», № 26 (60), с. 115–119.

Калдиева-Захаријева Ст., 2013: *Българска фразеология*. Sofia.

Караулов Ю.Н., *Русская языковая личность и задачи ее изучения*. В: *Язык и личность*. Ред. Д. Н. Шмелёв. Москва, с. 3–9.

Ковтун Л. С., 1975: *Лексикография в Московской Руси XVI – начала XVII в.* Ленинград.

Кожухова Л. В., 2007: *Речевой акт обращения*. «Вестник Ставропольского государственного университета», № 486, с. 80–85.

Козырев В. А., Черняк В. Д., 2000: *Вселенная в алфавитном порядке: Очерки о словарях русского языка*. Санкт-Петербург.

Колшанский Г. В., 1975: *Соотношение субъективных и объективных факторов в языке*. Москва.

Колшанский Г.В., 1980: *Контекстная семантика*. Москва.

Комиссаров В. Н., 2011: *Современное переводоведение*. Москва.

Краснов В. А., 2005: *Коммуникемы со значением "волеизъявления" в языке и речи: На материале русского языка*. Ростов-на-Дону.

Красных В., 2016: *Словарь и грамматика лингвокультуры. Основы психолингвокультурологии*. Москва.

Кубрякова Е. С., 1991: *Особенности речевой деятельности и проблемы внутреннего лексикона*. В: *Человеческий фактор в языке и речи: Язык и порождение речи*. Ред. Е. С. Кубрякова. Москва, с. 82–140.

Кузнецов С., 2000: *Большой толковый словарь русского языка*. Санкт-Петербург.

Ленько Г. Н., 2015: *Анализ категории эмотивности и смежных с нею понятий*. «Вестник Ленинградского государственного университета имени А. С. Пушкина», № 1, с. 84–91.

Лю Э., 2020: *Устойчивое выражение «честно говоря» как метатекстовый оператор*. «Мир науки, культуры, образования», № 1 (80), с. 341–342.

Львовская З. Д., 1985: *Теоретические проблемы перевода*. Москва.

Ляшевская О. Н., Плунгян В. А., Сичинава Д. В., 2025: *О морфологическом стандарте Национального корпуса русского языка*. «Вестник Московского университета. Серия 9: Филология», № 4, с. 111–135.

Малкова А. И.; Мирошниченко Л. Н.. 2018: *Лексикографические аспекты перевода как актуальная проблема современной лингвистики*. «Слово. Словарь. Термин. Лексикограф: сборник статей по материалам Международной научно-практической конференции памяти доктора филологических наук, профессора Ю. Н. Марчука», с. 419–423.

Мамонтова В.В., 2010: *Корпусная лингвистика в современной языковедческой парадигме*. «Актуальные вопросы современной науки», № 12, с. 230–238.

Маник С. А., 2019: *Параллельный корпус в практике перевода общественно-политических текстов (с английского на русский)*. «Modern Studies of Social Issues», vol. 11, № 4, с. 225–245.

Меликян В. Ю., 2001: *Об основных типах нечлененых предложений в русском языке*. «Филологические науки», № 6, с. 79–89.

Мокиенко В. М., 1980: *Славянская фразеология*. Москва.

Мокиенко В.М., Никитина Т.Г., 2007: *Большой словарь русских поговорок*. Москва.

Невских Н.Н., 2024: *Стратегия аудиовизуального перевода сериала «Шерлок» с английского языка на русский язык*. Томск.

Пермяков Г. Л., 1970: *От поговорки до сказки: Заметки по общей теории клише*. Москва.

Плунгян В. А., 2005: *Зачем нужен национальный корпус русского языка? В: Национальный корпус русского языка: 2003–2005. Результаты и перспективы*. Москва, с. 6-20.

Поливанов Е., 1968: *Статьи по общему языкознанию: Избранные работы*. Москва.

Резунова М. В., 2008: *Словарная статья: текст или дискурс?* «Известия Волгоградского государственного педагогического университета», с. 47-49.

Симонов П. В., 1981: *Эмоциональный мозг*. Москва.

Сичинава Д. В., 2012: *Parallel Corpora within the Russian National Corpus*. „Prace Filologiczne”, T. LXIII, s. 271–278.

Сичинава Д. В., 2013: *Параллельные корпуса в составе Национального корпуса русского языка как инструмент лексико-грамматической типологии*. В: *Tipología léxica*. Red. R. G. Tirado, I. A. Votyakova. Гранада, с. 71–78.

Сичинава Д. В., 2015: *Параллельные тексты в составе Национального корпуса русского языка: новые направления развития и результаты*. «Труды Института русского языка им. В. В. Виноградова», № 3 (6), с. 194–234.

Сичинава Д. В., 2019: *Параллельные тексты в составе Национального корпуса русского языка: новые языки и новые задачи*. «Труды Института русского языка им. В. В. Виноградова», Вып. 2, с. 41–60.

Солдатова О. С., 2009: *Использование фатических эмотивов в межкультурной коммуникации на примере ситуации «комплимент»*. «Вестник Российского университета дружбы народов», Серия: Лингвистика, № 4, с. 41–48.

Стахнова О. В., 2013: *К вопросу о фатической стороне речевого общения*. «Общественные науки. Языкознание», с. 108–115.

Тимофеева М. К., 2015: *Интроекция в лингвистику и в языке*. «Вопросы языкознания», № 6, с. 33–53.

Трофимов С. В., 2018: *Возможности использования параллельных корпусов в лингвистических исследованиях*. «Мировая наука», № 5(14), с. 383–387.

Трофимова Н. А., 2008: *Эмотивный смысл высказывания и операторы его порождения*. «Известия Российского государственного педагогического университета им. А. И. Герцена», с. 154–160.

Труфанова И. В., 2000: *Прагматика несобственно-прямой речи: Монография*. Москва.

Фёдоров А. И., 2008: *Фразеологический словарь русского литературного языка*. Москва.

Хныкиина Е. А., 2018: *Коммуникемы в разговорной речи студентов: экспериментальные данные*. «Актуальные проблемы филологии», № 17, с. 110–115.

Цой А., 2014: *Характеристика словарных статей в аспекте теории текста (на материале словарей служебных слов)*. W: *W kręgu zagadnień semantyki i stylistyki tekstu*. Red. A. Piasecka, I. Blumental. Łódź, s. 13–22.

Чернышова Т. В., 2004: *Фатическая речь как социальный символ коммуникации (на материале текстов печатных СМИ)*. «Известия Алтайского государственного университета», Серия: история, филология, философия и педагогика, № 4, с. 46–51.

Чибухашвили В. А., *Функциональные особенности языковой интуиции и условия ее использования в речевом развитии школьников*. «Вестник Томского государственного университета», № 4 (72), с. 63–69.

Шанский Н.М., 1957: *Лексика и фразеология современного русского языка*. Москва.

Шанский Н. М., 1996: *Фразеология современного русского языка*. Санкт-Петербург.

Шаров С.А., 2003: *Представительный корпус русского языка в контексте мирового опыта*. «Научно-техническая информация. Серия 2. Информационные процессы и системы», № 6, с. 9–17.

Шаховский В. И., 2009: *Эмоции как объект исследования в лингвистике*. «Вопросы психолингвистики», №. 9, с. 29–43.

Шингаров Г. Х., 1971: *Эмоции и чувства как формы отражения действительности*. Москва.

Щец Н., 2005: *Роль языка в структурировании знания*. Тверь.

Щерба Л. В., 1958: *Избранные работы по языкознанию и фонетике*. Ленинград.

Słowniki i encyklopedie

Akademický slovník současné češtiny, <https://www.slovníkcestiny.cz/uvod.php>.

- Basaj M., Siatkowski J., 1991: *Polsko-český slovník*. Warszawa.
- Čermák F., 1983: *Slovník české frazeologie a idiomatiky. 1. Přirovnání*. Praha.
- Čermák F., 1988: *Slovník české frazeologie a idiomatiky. 2. Výrazy neslovesné*. Praha.
- Čermák, F. 1994: *Slovník české frazeologie a idiomatiky. 3. Výrazy slovesné*. Praha.
- Čermák F., 2009: *Slovník české frazeologie a idiomatiky 4. Výrazy větné*. Praha.
- Čermák, F. 2023: *Slovník české frazeologie a idiomatiky. 5. Onomaziologický slovník*. Praha.
- dobryslownik.pl*, [https //dobryslownik.pl](https://dobryslownik.pl).
- dobryslownik.cz*, <http://www.dobryslownik.cz/cestina>.
- Hessen D., Stypuła R., 1967: *Большой польско-русский словарь*. Warszawa.
- Inny słownik języka polskiego*. Pod red. M. Bańko. Warszawa 2000.
- Nevzorova-Kmech E., 2010: *Polsko-rosyjski słownik frazeologiczny gwary młodzieżowej*. Łódź.
- Nový encyklopedický slovník češtiny*, <https://www.czechency.org>.
- Oliva K., 2012: *Polsko-český slovník*. Praha.
- Orłoś T. Z., 2010: *Wielki czesko-polski słownik frazeologiczny*. Kraków.
- Podręczny idiomatykon polsko-rosyjski (zeszyty 1–10)*. Red. W. Chlebda. Opole, 2006–2019.
- Skorupka S., 1967: *Słownik frazeologiczny języka polskiego*. Warszawa.
- Slovník současné češtiny LINGEA*, Brno 2011.
- Slovník současné češtiny Lingea*. Praha 2024.
- Slovník synonym*, <https://esynonyma.cz/>.
- Słownik języka polskiego PWN*, [https //sjp.pwn.pl](https://sjp.pwn.pl).
- Słownik języka polskiego PWN*. Pod red. M. Szymczak. Warszawa 1978–1981.
- Słownik Lingea*, <https://slowniki.lingea.pl/czesko-polski>.
- Słownik synonimów języka polskiego online*, <https://synonim.net/>.

Trávniček F., 1937: *Slovník jazyka českého*. Praha.

Uniwersalny słownik języka polskiego. Pod red. S. Dubisz. Warszawa 2003

Wielki słownik języka polskiego, <https://wsjp.pl>.

Большой академический словарь русского языка. Под ред. Л. И. Балахоновой. Москва–Санкт-Петербург 2013.

Картаслов, <https://kartaslov.ru>.

Малый русский словарь, ред. А. П. Евгеньевой. Москва 1999.

Словарь синонимов русского языка, <https://sinonim.org/>.

Словарь синонимов, <https://synonymonline.ru/>.

Толковый словарь русской разговорной речи, ред. Л. П. Крысина. Москва 2014–2022.

Тришин В. Н., 2013: *Словарь синонимов*. <https://rus-yaz.niv.ru/doc/synonyms-trishin/index.htm> (дата обращения: 25.02.2025)

Korpusy

Korpus Współczesnego Języka Polskiego, <https://kwjp.ipipan.waw.pl/>.

Narodowy Korpus Języka Czeskiego, <https://www.korpus.cz/>

Narodowy Korpus Języka Rosyjskiego, <https://ruscorpora.ru>

Narodowy Korpus Języka Polskiego, <https://nkjp.pl/>

Wyszukiwarka korpusowa MoncoPL, <https://monco-pl.clarin-pl.eu/>

Intercorp

Źródła internetowe

Burdík J., 2018: *Dělej si, co chceš!*

<https://www.svobodauceni.cz/clanek/delej-si-co-chces/> (datum přístupu: 06.05.2025).

Článek „5 běžných frází ve vztazích, které muži a ženy chápou jinak“.

<https://prozeny.blesk.cz/clanek/pro-zeny-laska-a-vztahy/600034/5-beznych-frazi-ve-vztazich-ktre-muzi-a-zeny-chapou-jinak.html#1310130> (datum přístupu: 06.05.2025).

Glosbe, <https://pl.glosbe.com/>

Google Translate,

<https://translate.google.pl>.

Jak berete, když vám partner/ka řekne „jak chceš“?

<https://www.emimino.cz/diskuse/jak-chces-482056/strankovani/2/> (datum přístupu: 06.05.2025).

Jelínek T., Korpus SYN2020: Automatická anotace.

[https://wiki.korpus.cz/doku.php/cnk:syn2020:automaticka_anotace?s\[\]=morfologick%C3%A9&s\[\]=zna%C4%8Dkov%C3%A1n%C3%AD](https://wiki.korpus.cz/doku.php/cnk:syn2020:automaticka_anotace?s[]=morfologick%C3%A9&s[]=zna%C4%8Dkov%C3%A1n%C3%AD) (datum přístupu: 06.05.2025).

Pragmaticon.

<https://pragmaticon.ruscorpora.ru/> (дата обращения: 03.02.2025).

Pravopisně.cz. Popravdě-po pravdě.

https://www.pravopisne.cz/2015/08/popravde-x-po-pravde/#google_vignette (datum přístupu: 10.05.2025).

Wielki Słownik Języka Polskiego PAN. Hasło „Bez dyskusji”.

<https://wsjp.pl/haslo/podglad/40018/bez-dyskusji> (data dostępu: 26.02.2025).

Ženský česko-český výkladový slovník.

https://www.fzu.cz/~sipr/documents/zensky_muzsky_slovník.html (datum přístupu: 06.05.2025).

7 фраз, которые говорят лжецы.

<https://heroine.ru/7-fraz-kotorye-govoryat-lzheczy-chtoby-zastavit-poverit-v-ih-slova/?ysclid=mdbfbnk8rw690266048> (дата обращения: 10.05.2025).

Захаров В. П., 2014: Корпусная лингвистика в России. В: Материалы международной конференции CrossLingua '2014. Санкт-Петербург.

http://crosslingua.cfuv.ru/publications/2014_2_zakharov.pdf (дата обращения: 05.04.2024).

Значение словосочетания «делать своё дело».

kartaslov.ru/значение-слова/делать+своё+дело (дата обращения: 05.05.2025).

Значение фразы «если честно».

<https://calc-kirpich.ru/stat/polezno/znacenie-frazy-esli-cestno?ysclid=mdbfcrhlps377827177> (дата обращения: 10.05.2025).

Грамота.ру. Откровенно говоря.

[gramota.ru/poisk?query=откровенно говоря&mode=slovari&ysclid=mdbfedw060601347209](http://gramota.ru/poisk?query=откровенно%20говоря&mode=slovari&ysclid=mdbfedw060601347209) (дата обращения: 10.05.2025).

Грамота.ру. По правде говоря/сказать.

<https://gramota.ru/biblioteka/spravochniki/spravochnik-po-punktuatsii/po-pravde-govorya-skazat?> (дата обращения: 10.05.2025).

Манипуляция «делай что хочешь» в отношениях.

<https://dzen.ru/a/ZGB54lEDcnnUTguL> (дата обращения: 05.05.2025).

Наумов Е., Делай что угодно.

<https://syg.ma/@affabricatus/dielai-hto-ughodno-mnoghotochiie> (дата обращения: 5.05.2025).

Программа MyStem.

<https://tech.yandex.ru/mystem/> (дата обращения: 18.12.2023).

Савкин О., 2024: Что на самом деле значит фраза «делай, что хочешь»: скрытые смыслы и послания.

<https://www.psychologies.ru/articles/hto-na-samom-dele-znachit-fraza-delai-hto-khochesh-skrytye-smysly-i-poslaniya/> (дата обращения: 05.05.2025).

Учебный фразеологический словарь. Вот и всё.

https://phraseologiya.academic.ru/171/вот_и_всё (дата обращения: 26.02.2025).

Учебный фразеологический словарь. По правде говоря.

phraseologiya.academic.ru/805/по_правде_говоря? (дата обращения: 10.05.2025).

Учебный фразеологический словарь. Честно говоря.

phraseologiya.academic.ru/1127/честно_говоря (дата обращения: 10.05.2025).

Что значит «делать своё дело?»

<https://otvet.mail.ru/question/166912763?utm> (дата обращения: 05.05.2025).

Что значит «ты что спятил?»

<https://otvet.mail.ru/question/176885078> (дата обращения: 20.06.2025).

Что значит выражение «Честно говоря»?

<https://otvet.mail.ru/question/231635800> (дата обращения: 05.05.2025).

Что значит фраза как хочешь?

<https://otvet.mail.ru/question/64636040?utm> (дата обращения: 05.05.2025).

Что такое откровенно говоря?

<https://telegra.ph/CHto-takoe-otkrovenno-govorya-12-24?ysclid=mdbfhlcw2c579880075> (дата обращения: 10.05.2025).

STRESZCZENIE

Celem rozprawy jest opracowanie metod wyodrębniania, analizy i opisu wielowyrazowych struktur języka mówionego o funkcji fatycznej i ematywnej w języku polskim, rosyjskim i czeskim. Punktem odniesienia stała się obserwacja, że tradycyjna leksykografia – zarówno jednojęzyczna, jak i przekładowa – w ograniczonym stopniu rejestruje formy złożone charakterystyczne dla komunikacji ustnej, ponieważ koncentruje się na pojedynczym wyrazie jako podstawowej jednostce opisu słownikowego.

Problem ten, obecny od dawna w badaniach frazeologicznych i leksykograficznych, wynika z trudności związanych z wyznaczeniem granic takich konstrukcji, ich wariantowością, wieloznacznością oraz marginalizacją języka mówionego. W pracy za jednostkę wielowyrazową uznaje się *reprodukt*, rozumiany – zgodnie z ujęciem Wojciecha Chlebdy – jako utrwalona jednostka języka, wyodrębniana z tekstów na podstawie jej regularnej odtwarzalności i pełniąca funkcję werbalizatora intencji nadawcy. Rozprawa podejmuje próbę wypełnienia tej luki poprzez opis i klasyfikację reproduktów fatycznych i ematywnych oraz próbę ustalenia ich ekwiwalentów przekładowych z wykorzystaniem zasobów korpusowych i pozakorpusowych.

Materiał badawczy został głównie pozyskany z NKJP, NKJR, NKJCz, korpusu MoncoPL oraz korpusu równoległego InterCorp. Zastosowanie tych narzędzi umożliwiło empiryczne wyodrębnienie jednostek wielowyrazowych w autentycznych kontekstach oraz ich porównanie w trzech językach. Szczególną rolę w tym procesie odegrały korpusy równoległe, pozwalające na badanie ekwiwalentów tłumaczeniowych z uwzględnieniem kryterium sytuacyjności i funkcji pragmatycznej. Jako uzupełnienie wykorzystano źródła pozakorpusowe, takie jak słowniki jednojęzyczne i przekładowe oraz materiały pochodzące z internetu (m.in. fora i portale językowe). W pracy przyjęto kryterium, że korpus stanowi podstawę badań i narzędzie wspomagające proces analizy, jednak nie pełni funkcji instancji rozstrzygającej. Dane korpusowe wymagają konfrontacji z innymi źródłami i interpretacji w szerszym kontekście leksykograficznym.

Rozprawa ma charakter teoretyczno-analityczny.

Część pierwsza (rozdziały 1–5) zawiera zarys teoretyczny, omówienie narzędzi badawczych, stan badań nad leksykografią jednojęzyczną i przekładową w Polsce, Rosji i Czechach oraz podstawowe założenia dotyczące rozumienia jednostki wielowyrazowej, reprodukowalności i lingwistyki korpusowej.

Część druga (rozdziały 6–8) ma charakter badawczy – przedstawiono w niej metody wyodrębniania wielowyrazowych jednostek języka mówionego (m.in. metodę słownikowo-kolokacyjną, synonimiczną i tłumaczeniową) oraz metody ustalania ich ekwiwalentów tłumaczeniowych z wykorzystaniem korpusów równoległych czy materiału audiowizualnego.

Zwieńczeniem pracy jest koncepcja autorskiego słownika polsko-rosyjskiego i polsko-czeskiego, zawierająca propozycję makrostruktury i opisu artykułu hasłowego. Opracowanie to stanowi efekt połączenia rozwiązań wypracowanych w dotychczasowej leksykografii z autorskimi, nowatorskimi ujęciami, dostosowanymi do specyfiki języka mówionego i możliwości opracowania elektronicznego. Rezultaty przeprowadzonych badań potwierdzają, że język mówiony, mimo iż bywa marginalizowany w tradycyjnych opisach leksykograficznych, stanowi cenne źródło utrwalonych i regularnie odtwarzanych struktur wielowyrazowych, będących nieodłącznym elementem codziennej komunikacji. Wykorzystanie metod korpusowych umożliwiło identyfikację tego typu jednostek oraz wskazanie ich potencjalnych odpowiedników przekładowych. Należy jednak zauważyć, że zasoby tekstów mówionych w korpusach są ograniczone, a część z nich stanowią materiały stylizowane na język ustny. Mimo to korpusy niezmiennie pozostają wartościowym narzędziem wspomagającym proces poszukiwania i weryfikacji jednostek, umożliwiającym formułowanie wstępnych hipotez dotyczących ich funkcjonowania i ekwiwalencji w innych językach.

Rozprawa porusza wiele zagadnień leksykografii przekładowej, frazematyki i lingwistyki korpusowej, proponując nowy model opisu wielowyrazowych jednostek języka w ujęciu międzyjęzykowym i translatorycznym.

SUMMARY

The aim of this dissertation is to develop methods for isolating, analyzing, and describing multi-word structures of spoken language with phatic and emotive functions in Polish, Russian, and Czech. The point of reference is the observation that traditional lexicography – both monolingual and bilingual – records complex forms characteristic of oral communication only to a limited extent, as it focuses primarily on the single word as the basic unit of lexicographic description.

This issue, long present in phraseological and lexicographic research, arises from the difficulties involved in defining the boundaries of such constructions, their variability, ambiguity, and the marginalisation of spoken language. In this dissertation, a multi-word unit is understood as a *reproduct*—following the definition proposed by Wojciech Chlebda—namely, a fixed linguistic unit extracted from texts on the basis of its regular reproducibility and functioning as a verbaliser of the speaker’s communicative intention. The dissertation seeks to bridge this gap by describing and classifying phatic and emotive reproducts and by attempting to determine their translational equivalents using both corpus-based and non-corpus resources.

The research material was primarily drawn from the Polish National Corpus, the Russian National Corpus, the Czech National Corpus, the MoncoPL corpus, and the parallel corpus InterCorp. The use of these tools made it possible to empirically extract multi-word units in authentic contexts and to compare them across the three languages. A particularly important role in this process was played by parallel corpora, which made it possible to examine translational equivalents with regard to situational and pragmatic criteria. As a complement, non-corpus sources were also used, such as monolingual and bilingual dictionaries, as well as online materials (including discussion forums and language portals). The study adopts the principle that the corpus serves as the foundation of the research and as a tool supporting the analytical process, but it does not function as the ultimate authority. Corpus data require confrontation with other sources and interpretation within a broader lexicographic context.

The dissertation is of a theoretical and analytical nature.

The first part (Chapters 1–5) provides a theoretical overview, a discussion of research tools, a review of studies on monolingual and bilingual lexicography in Poland, Russia, and the Czech Republic, as well as the fundamental assumptions concerning the understanding of multi-word units, reproducibility, and corpus linguistics.

The second part (Chapters 6–8) is of an empirical nature and presents methods for extracting multi-word units of spoken language (including the dictionary–collocational, synonymic, and translational methods), as well as methods for determining their translational equivalents using parallel corpora and audiovisual material.

The culmination of the dissertation is the concept of an original Polish–Russian and Polish–Czech dictionary, which includes a proposal for its macrostructure and the design of the dictionary entry. This work is the result of combining solutions developed in previous lexicographic practice with original and innovative approaches adapted to the specificity of spoken language and the possibilities of electronic processing. The results of the conducted research confirm that spoken language, although often marginalised in traditional lexicographic descriptions, constitutes a valuable source of fixed and regularly reproduced multi-word structures that form an integral part of everyday communication. The application of corpus-based methods made it possible to identify such units and to indicate their potential translational equivalents. It should be noted, however, that the resources of spoken texts in corpora are limited, and some of them consist of materials stylised as oral language. Nevertheless, corpora consistently remain a valuable tool supporting the process of identifying and verifying units, enabling the formulation of preliminary hypotheses concerning their functioning and equivalence in other languages.

The dissertation addresses a wide range of issues in translational lexicography, phrasematics, and corpus linguistics, proposing a new model for describing multi-word language units in an interlingual and translational perspective.

РЕЗЮМЕ

Цель данной диссертации заключается в разработке методов извлечения, анализа и описания многолексемных структур устной речи, выполняющих фатические и эмотивные функции в польском, русском и чешском языках. Отправной точкой послужило наблюдение, что традиционная лексикография — как одноязычная, так и переводная — в ограниченной степени фиксирует сложные формы, характерные для устной коммуникации, поскольку сосредотачивается на отдельном слове как основной единице словарного описания.

Эта проблема, существующая уже давно в фразеологических и лексикографических исследованиях, обусловлена трудностями, связанными с определением границ подобных конструкций, их вариантностью, многозначностью, а также маргинализацией разговорной речи. В настоящей работе за многолексемную единицу принимается *репродукт*, понимаемый — в соответствии с концепцией Войцеха Хлебды — как закреплённая в языке единица, выделяемая из текстов на основании её регулярной воспроизводимости и выполняющая функцию вербализатора интенции говорящего. Диссертация представляет собой попытку восполнить данный пробел посредством описания и классификации фатических и эмотивных репродуктов, а также попытку установления их переводческих эквивалентов с использованием корпусных и внекорпусных источников.

Материалы исследования были получены в основном из Национального корпуса польского языка, Национального корпуса русского языка, Национального корпуса чешского языка, корпуса MoncoPL и параллельного корпуса InterCorp. Использование этих инструментов позволило эмпирически выделить многолексемные единицы в аутентичных контекстах и провести их сопоставление в трёх языках. Особую роль в этом процессе сыграли параллельные корпуса, которые обеспечили возможность исследования переводческих эквивалентов с учётом критерия ситуативности и прагматической функции. В качестве дополнения были использованы внекорпусные источники, такие как одноязычные и переводные словари, а также материалы из интернета (в частности, языковые порталы и форумы). В работе принято положение, согласно которому корпус рассматривается как основа исследования и инструмент, поддерживающий аналитический процесс, однако он не выполняет функции

окончательной инстанции. Корпусные данные требуют сопоставления с другими источниками и интерпретации в более широком лексикографическом контексте.

Диссертация носит теоретико-аналитический характер.

Первая часть (главы 1–5) содержит теоретический обзор, обсуждение исследовательских инструментов, состояние исследований в области одноязычной и переводной лексикографии в Польше, России и Чехии, а также основные предположения относительно понимания многолексемных единиц, воспроизводимости и корпусной лингвистики.

Вторая часть (главы 6–8) носит исследовательский характер — в ней представлены методы выделения многолексемных единиц устной речи (в частности, словарно-коллокационный, синонимический и переводческий методы), а также методы установления их переводческих эквивалентов с использованием параллельных корпусов и аудиовизуального материала.

Завершением работы является концепция авторского польско-русского и польско-чешского словаря, включающая предложение макроструктуры и описание словарной статьи. Данное исследование представляет собой результат сочетания решений, выработанных в предыдущей лексикографической традиции, с авторскими инновационными подходами, адаптированными к специфике устной речи и возможностям электронного представления. Результаты проведённого исследования подтверждают, что устная речь, хотя и нередко маргинализируется в традиционных лексикографических описаниях, представляет собой ценнейший источник устойчивых и регулярно воспроизводимых многолексемных структур, являющихся неотъемлемым элементом повседневной коммуникации. Использование корпусных методов позволило идентифицировать данные типы единиц и указать их потенциальные переводческие эквиваленты. Однако следует отметить, что объём устных текстов в корпусах ограничен, и часть из них представляет собой материалы, стилизованные под устную речь. Тем не менее корпуса по-прежнему остаются ценным инструментом, поддерживающим процесс поиска и верификации единиц, а также позволяющим формулировать предварительные гипотезы относительно их функционирования и эквивалентности в других языках.

Диссертация затрагивает широкий круг вопросов переводной лексикографии, фразематики и корпусной лингвистики, предлагая новую модель описания многолексемных языковых единиц в межъязыковом и переводческом аспектах.

SHRNUTÍ

Cílem disertační práce je vypracování metod pro izolaci, analýzu a popis víceslovných struktur mluveného jazyka s fatickou a emotivní funkcí v polštině, ruštině a češtině. Východiskem se stalo zjištění, že tradiční lexikografie – jak jednojazyčná, tak překladová – pouze v omezené míře zaznamenává víceslovné formy charakteristické pro ústní komunikaci, neboť se soustředí na jednotlivé slovo jako na základní jednotku lexikografického popisu.

Tento problém, přítomný již dlouhou dobu ve frazeologických a lexikografických výzkumech, vyplývá z obtíží spojených s vymezením hranic takových konstrukcí, jejich variantností, mnohoznačností a marginalizací mluveného jazyka. V této práci je za víceslovnou jednotku považován *reprodukt*, chápaný – v souladu s pojetím Wojciecha Chlebdy – jako ustálená jazyková jednotka, vyčleňovaná z textů na základě její pravidelné reprodukovatelnosti a plnící funkci verbalizátoru intence mluvčího. Disertační práce se pokouší zaplnit tuto mezeru prostřednictvím popisu a klasifikace fatických a emotivních reproduktů a také prostřednictvím snahy o určení jejich překladových ekvivalentů s využitím korpusových i mimokorpusových zdrojů.

Výzkumný materiál byl získán především z Národního korpusu polského jazyka, Národního korpusu ruského jazyka, Národního korpusu českého jazyka, korpusu MoncoPL a z paralelního korpusu InterCorp. Využití těchto nástrojů umožnilo empirické vyčlenění víceslovných jednotek v autentických kontextech a jejich srovnání ve třech jazycích. Zvláštní roli v tomto procesu sehrály paralelní korpusy, které umožnily zkoumání překladových ekvivalentů s ohledem na kritérium situačnosti a pragmatické funkce. Jako doplněk byly využity mimokorpusové zdroje, jako například jednojazyčné a překladové slovníky a také materiály pocházející z internetu (např. jazyková fóra a portály). V práci bylo přijato východisko, že korpus představuje základ výzkumu a nástroj podporující analytický proces, avšak neplní funkci konečné autority. Korpusová data vyžadují konfrontaci s jinými zdroji a interpretaci v širším lexikografickém kontextu.

Disertační práce má teoreticko-analytický charakter.

První část (kapitoly 1–5) zahrnuje teoretický přehled, popis výzkumných nástrojů, stav zkoumání jednojazyčné a překladové lexikografie v Polsku, Rusku a Česku a také základní východiska týkající se chápání víceslovné jednotky, reprodukovatelnosti a korpusové lingvistiky.

Druhá část (kapitoly 6–8) má výzkumný charakter – jsou v ní představeny metody vyčleňování víceslovných jednotek mluveného jazyka (mj. metoda slovníkově-kolokační, synonymická a překladová) a také metody určování jejich překladových ekvivalentů s využitím paralelních korpusů a audiovizuálního materiálu.

Závěrem práce je koncepce autorského polsko-ruského a polsko-českého slovníku, která zahrnuje návrh makrostruktury a popis slovníkového hesla. Toto zpracování představuje výsledek propojení řešení vypracovaných v dosavadní lexikografii s autorskými, inovativními přístupy přizpůsobenými specifice mluveného jazyka a možnostem elektronického zpracování. Výsledky provedeného výzkumu potvrzují, že mluvený jazyk, ačkoli bývá v tradičních lexikografických popisech často marginalizován, představuje cenný zdroj ustálených a pravidelně reprodukováných víceslovných struktur, které tvoří nedílnou součást každodenní komunikace. Využití korpusových metod umožnilo identifikaci tohoto typu jednotek a určení jejich potenciálních překladových ekvivalentů. Je však třeba poznamenat, že rozsah mluvených textů v korpusech je omezený a část z nich tvoří materiály stylizované na mluvený jazyk. Přesto korpusy i nadále představují cenný nástroj podporující proces vyhledávání a ověřování jednotek, který umožňuje formulovat předběžné hypotézy o jejich fungování a ekvivalenci v jiných jazycích.

Disertační práce se zabývá řadou otázek překladové lexikografie, frazematiky a korpusové lingvistiky a zároveň navrhuje nový model popisu víceslovných jazykových jednotek v mezijazykovém a translačním pojetí.