

POLSKIE TOWARZYSTWO RUSYCYSTYCZNE

przeгляд rusycystyczny

2022, nr 1 (177)

Katowice 2022

KOMITET HONOROWY

Janusz Henzel, Walenty Piłat, Władysław Woźniewicz

KOMITET REDAKCYJNY

Tadeusz Klimowicz — przewodniczący

Franciszek Apanowicz

Petar Bunjak

Jens Herlth

Jewgienij Jabłokow

Władimir Klimonow

Joanna Madloch

Daria Nevszkaya

Kadisha Nurgali

Antoaneta Olteanu

Grzegorz Przebinda

Barbara Stempczyńska

Walerij Tiupa

Halina Waszkielewicz

ZESPÓŁ REDAKCYJNY

Piotr Fast (redaktor naczelny), Michał Głuszkowski (zastępca redaktora naczelnego), Justyna Pisarska, Joanna Darda-Gramatyka, Anna Paszkowska (sekretarz redakcji)

Adiustacja tekstów obcojęzycznych

Yevheniy Liashchevskiyi

Piotr Plichta

Korekta

Justyna Pisarska

ADRES REDAKCJI

Przegląd Ruscystyczny, 41-500 Chorzów, ul. Zjednoczenia 1/11

tel. +48 604 965 737; +48 505 300 667

e-mail: prz.rus@op.pl

<http://www.journals.us.edu.pl/index.php/PR>

Index 371866

ISSN 0137-298X

**Polskie
Towarzystwo
Ruscystyczne**

Polskie Towarzystwo Ruscystyczne
41-200 Sosnowiec
ul. Stefana Grota Roweckiego 5/318
tel. 505 300 667, ptr.polska@gmail.com
www.peteer.pl

WYDAWCY



**UNIWERSYTET ŚLĄSKI
W KATOWICACH**

Uniwersytet Śląski
Instytut Literaturoznawstwa
Instytut Językoznawstwa
41-200 Sosnowiec
ul. Stefana Grota Roweckiego 5

Polskie Towarzystwo Rusycystyczne, Kolegium Redakcyjne „Przeglądu Rusycystycznego”, redakcja naszego pisma, współpracownicy i autorzy z głębokim smutkiem i wstydem przyjęli decyzję władz Federacji Rosyjskiej o rozpoczęciu działań wojennych przeciwko Ukrainie oraz wkroczeniu wojsk rosyjskich na teren tego niezawisłego państwa zamieszkiwanego przez dumny naród, który niedawno uzyskał należną mu państwowość. Ufamy, że nie wszyscy znaczący politycy rosyjscy solidaryzują się z oficjalnymi decyzjami prezydenta i że zrobią, co w ich mocy, by powstrzymać agresję Rosji na Ukrainie. Jesteśmy z Ukrainą — dzisiaj wszyscy jesteśmy Ukraińcami! Protestujemy przeciwko działaniom, w których giną niewinni ludzie. Obawiający się o utratę władzy rosyjski dyktator pokrętnymi sposobami zmierza do opanowania umysłów Rosjan przez pseudopatriotyczną propagandę, rozbudzającą imperialne roszczenia. Wojna przeciwko Ukrainie jest działaniem ludobójczym!

Z satysfakcją przyjmujemy wieści o protestach środowisk naukowych i kulturalnych Rosji. Tysiące tego rodzaju oświadczeń najwybitniejszych Rosjan znaleźć można na przykład na portalach internetowych: <https://newizv.ru/news/society/26-02-2022/dve-tysyachi-deyateley-kultury-potrebovali-prekratit-voynu-s-ukrainoy>; Российские шекспироведы выступили против войны с Украиной: philologist — LiveJournal; <https://echo.msk.ru/blog/echomsk/2985101-echo/>.

Są wśród nich Jurij Apresjan, Julij Kim, Ludmiła Ulicka, Dmitrij Głuchowski, Wiktor Jerofiejew, Boris Akunin, Ludmiła Pietruszewska, Dmitrij Bykow, Andriej Zwiagincew, Lew Rubinstein, Władimir Sorokin, Guzel Jachina, Wiktor Szenderowicz, Nikołaj Swanidze, Jurij Domanskij i wielu, wielu innych. Potępiają oni decyzję władz i nawołują do rozsądku. Przyłączamy się do nich, mając nadzieję, że nasz głos, choćby cichy i mało znaczący, wesprze tych, którzy starają się uniemożliwić lub choćby utrudnić realizację zbrodniczych celów rosyjskiego reżimu.

Zarząd Polskiego Towarzystwa Rusycystycznego
Kolegium Redakcyjne i Redakcja „Przeglądu Rusycystycznego”

Польское русистское общество, редакционный комитет журнала «Русское обозрение», редакция нашего журнала, сотрудники и авторы с глубоким сожалением и чувством стыда восприняли решение властей Российской Федерации о начале военных действий против Украины и вторжении российских войск на территорию этого независимого государства, населенного гордым народом; государства, не так давно обретшего положенную ему государственность. Мы верим, что не все авторитетные российские политики солидарны с официальными решениями президента, и что они сделают все от них возможное, чтобы не допустить эскалации военной истерии в России. Мы с Украиной – сегодня мы все украинцы! Мы протестуем против действий, которые приводят к гибели невинных людей. Российский диктатор, в страхе потерять власть, извращенными методами псевдопатриотической пропаганды, возрождающей имперские притязания, пытается контролировать умы российского народа. Война против Украины – это геноцид!

С чувством удовлетворения мы встречаем новости о протестах научных и культурных сообществ России (тысячи подобных заявлений известных россиян можно найти, например, на сайтах: <https://newizv.ru/news/society/26-02-2022/dve-tysyachi-deyateley-kultury-potrebovali-prekratit-voynu-s-ukrainoy>, <https://philologist.livejournal.com/12299331.html>, <https://echo.msk.ru/blog/echomsk/2985101-echo/>). В их числе заявления Юрия Апресяна, Юлия Кима, Людмилы Улицкой, Дмитрия Глуховского, Виктора Ерофеева, Бориса Акунина, Людмилы Петрушевской, Дмитрия Быкова, Андрея Звягинцева, Льва Рубинштейна, Владимира Сорокина, Гузель Яхины, Виктора Шендеровича, Николая Сванидзе и многих, многих других. Все они осуждают решение властей и призывают к здравому смыслу. Мы присоединяемся к ним с надеждой, что наш голос, пусть тихий и незначительный, поддержит тех, кто пытается предотвратить или хотя бы воспрепятствовать реализации преступных целей российского режима.

Руководство Польского русистического общества
Редакционный комитет и редакция журнала
«Русское обозрение»

Polskie Towarzystwo Rusycystyczne (The Polish Russian Studies Society), the Editorial Board of Przegląd Rusycystyczny (The Russian Studies Review) as well as the editorial team and authors and other contributors of the journal deeply condemn and deplore the decision of the authorities of the Russian Federation to initiate hostilities against Ukraine and the consequent encroachment of Russian troops on the territory of this independent state, inhabited by a proud nation which has recently obtained its due statehood. We trust that not all prominent Russian politicians approve and support the official decisions of the current president of the Russian Federation and that they will do what they can to stop Russia's aggression against Ukraine. We stand with Ukraine — today we are all Ukrainians! We fiercely protest against all actions in which innocent people are killed while the Russian dictator, fearful of losing power, is using twisted means to control the minds of Russian citizens through pseudo-patriotic propaganda that arouses imperial claims. The war against Ukraine is vile and genocidal!

We are pleased to hear about the protests coming from scientific and cultural circles of Russia as thousands of such statements by the most prominent Russians can be found, for example, on the Internet portals: <https://newizv.ru/news/society/26-02-2022/dve-tysyachi-deyatelye-kultury-potrebovali-prekratit-voynu-s-ukrainoy>; <https://philologist.livejournal.com/12299331.html>; <https://echo.msk.ru/blog/echomsk/2985101-echo/>.

The list of those who expressed their protest includes Yuri Apresyan, Yuli Kim, Lyudmila Ulicka, Dmitry Glukhovsky, Viktor Yerofeyev, Boris Akunin, Lyudmila Petrushevskaya, Dmitry Bykov, Andrei Zvyagintsev, Lev Rubinstein, Vladimir Sorokin, Guzel Yakhina, Victor Shenderovich, Nikolay Svanidze, Yuri Domansky and many others. They all condemn the decision made by the authorities of the Russian Federation and call for reason. We join their cry of protest in the hope that our voice, however quiet and insignificant, will support those who try to prevent or at least hinder the realization of the inhumane and criminal goals of the current Russian regime.

Board of Directors of Polskie Towarzystwo Rusycystyczne
(Polish Russian Studies Association)
Editorial Board and Team of Przegląd Rusycystyczny
(The Russian Studies Review)

SPIS TREŚCI

KORPUSY W BADANIACH NAD JĘZYKIEM

GOŚCINNIE POD REDAKCJĄ
ANDRZEJA CHARCIARKA

- 11 ANDRZEJ CHARCIAREK Wstęp
- 14 EWA BIAŁEK Polifunkcyjność korpusu językowego (na przykładzie platformy Sketch Engine i Narodowego Korpusu Języka Rosyjskiego)
- 32 DANIEL BORYSOWSKI Wykorzystanie korpusów rosyjskojęzycznych newsów internetowych na potrzeby systemów automatycznego rozpoznawania mowy w obszarze monitoringu mediów
- 55 ANDRZEJ CHARCIAREK Korpus równoległy a ekwiwalentyzacja
- 76 DARIA DAYTER Badania korpusowe nad tłumaczeniem pisemnym i ustnym
 ŁUKASZ GRABOWSKI na materiale tekstów w języku rosyjskim: przeszłość, teraźniejszość i przyszłość
- 98 ALEKSANDER KIKLEWICZ Badania korpusowe w składni: możliwości i ograniczenia
- 115 MACIEJ LABOCHA Korpusy wobec poszukiwania optymalnych kolokacji do celów leksykografii przekładowej
- 133 MARIA MOCARZ-KLEINDIENST Korpusy językowe w polsko-rosyjskim projekcie leksykograficznym *Architektura i urbanistyka*
- 153 ROMAN ROSZKO Rusycystyczne zasoby i narzędzia CLARIN-PL
- 177 TOMASZ SZUTKOWSKI W poszukiwaniu kolokacji i ich odpowiedników przekładowych z wykorzystaniem narzędzi elektronicznych i danych korpusowych
- 195 ALEKSANDRA ZIĘBA Korpus równoległy InterCorp w procesie ustalania polsko-rosyjskich par metaoperatorów
- * * *
- 215 YEVHEN LIASHCHEVSKYI Zainteresowanie Indiami i starożytnymi tekstami indyjskimi w literaturze rosyjskiej od końca XVIII w. do połowy drugiej dekady XX w.
- 224 GRZEGORZ PRZEBINDA Trzy spotkania z Andriejem Siniawskim (Paryż — Genewa — Kraków, 1987–1993)
- 249 ANDRZEJ ROMANOWSKI Gdy państwo depcze prawo...
 GRZEGORZ PRZEBINDA Wywiad z Marią Rozanową i Andriejem Siniawskim (listopad 1993 roku)

RECENZJE

- 261 ALEKSANDRA SZYMAŃSKA Валентина Е. Головчинер, Татьяна Л. Веснина, *Комическое в пьесах М. Булгакова 1920-х годов*, Издательство Томского университета, Томск 2021
- 264 GRAŻYNA MAROSZCZUK *Gustaw Herling-Grudziński. Między Wschodem a Zachodem*, red. Grzegorz Przebinda, Bartosz Gołąbek, Wojciech Gruchała, Wydawnictwo Humanistyczne Pigionianum, Karpacka Państwowa Uczelnia w Krośnie, Krosno 2021
- 275 JELENA WAŻENINA ROSYJSKA FRAZEOLOGIA JAKO LOS
ANATOLIJ ZAHNITKO Анатолий Жуков, *Очерки по фразеологической семантике: монография, 2-е издание, переработанное и дополненное*, Издательство РУСАЙНС, Москва 2020
- 283 Noty o autorach

СОДЕРЖАНИЕ

КОРПУСЫ ТЕКСТОВ В ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

ПОД РЕДАКЦИЕЙ АНДЖЕЯ ХАРЦЯРЕКА

- 11 АНДЖЕЙ ХАРЦЯРЕК Вступление
- 14 ЭВА БЯЛЭК Полифункциональность языкового корпуса (на примере программы Sketch Engine и Национального корпуса русского языка)
- 32 ДАНЕЛЬ БОРЫСОВСКИ Использование русскоязычных интернет-корпусов новостей для систем автоматического распознавания речи в сфере медиа-мониторинга
- 55 АНДЖЕЙ ХАРЦЯРЕК Параллельный корпус и эквивалентность
- 76 ДАРЬЯ ДАЙТЕР Корпусные исследования письменного и устного перевода
ЛУКАШ ГРАБОВСКИ на материале текстов на русском языке: прошлое, настоящее и будущее
- 98 АЛЕКСАНДР КИКЛЕВИЧ Корпусные исследования в синтаксисе: возможности и ограничения
- 115 МАЧЕЙ ЛЯБОХА Корпусы и поиски оптимальных коллокаций для целей переводной лексикографии
- 133 МАРИЯ МОЦАЖ-КЛЕЙДИЕНСТ Языковые корпуса в польско-русском лексикографическом проекте *Архитектура и градостроительство*
- 153 РОМАН РОШКО Русскоязычные ресурсы и инструменты на сайте CLARIN-PL
- 177 ТОМАШ ШУТКОВСКИ В поисках коллокаций и их переводческих эквивалентов с использованием электронных средств и корпусных данных
- 195 АЛЕКСАНДРА ЗЕМБА Параллельный корпус InterCorp в процессе установления польско-русских пар метаорепаторов

* * *

215	ЕВГЕН ЛЯЩЕВСКИЙ	Интерес к Индии и древнеиндийским текстам в русской литературе конца XVIII — середины XX века
224	ГЖЕГОЖ ПШЕБИНДА	Три встерчи с Андреем Синявским (Париж — Женева — Краков, 1987–1993)
249	АНДЖЕЙ РОМАНОВСКИ ГЖЕГОЖ ПШЕБИНДА	Если государство преступает закон... Интервью с Марией Розановой и Андреем Синявским (ноябрь 1993 г.)
РЕЦЕНЗИИ		
261	АЛЕКСАНДРА ШИМАНЬСКА	Валентина Е. Головчинер, Татьяна Л. Веснина, <i>Комическое в пьесах М. Булгакова 1920-х годов</i> , Издательство Томского университета, Томск 2021
264	ГРАЖИНА МАРОЩУК	<i>Gustaw Herling-Grudziński. Między Wschodem a Zachodem</i> , ред. Grzegorz Przebinda, Bartosz Gołąbek, Wojciech Gruchała, Wydawnictwo Humanistyczne Pigionium, Karpacka Państwowa Uczelnia w Krośnie, Krosn 2021
275	ЕЛЕНА ВАЖЕНИНА АНАТОЛИЙ ЗАГНИТКО	Анатолий Жуков, <i>Очерки по фразеологической семантике: монография</i> , 2-е издание, переработанное и дополненное, Издательство РУСАЙНС, Москва 2020
283	СВЕДЕНИЯ ОБ АВТОРАХ	

TABLE OF CONTENTS

CORPORA IN LANGUAGE RESEARCH

GUEST EDITOR

ANDRZEJ CHARCIAREK

11	ANDRZEJ CHARCIAREK	Preface
14	EWA BIAŁEK	Polyfunctionality of language corpus (on the material of the Sketch Engine and the Russian national corpus)
32	DANIEL BORYSOWSKI	The use of Russian-language internet news corpora for the purposes of automatic speech recognition systems in the area of the media monitoring
55	ANDRZEJ CHARCIAREK	Parallel corpus and equivalencing
76	DARIA DAYTER ŁUKASZ GRABOWSKI	Corpus-based and corpus-driven research on translation and interpreting in Russian: the past, the present and the future
98	ALEKSANDER KIKLEWICZ	Corpus research in syntax: opportunities and limitations
115	MACIEJ LABOCHA	Using corpora in search of optimal collocations for translation dictionaries

- 133 MARIA MOCARZ-KLEINDIENST Language corpora in a Polish-Russian lexicographical project *Architecture and urban planning*
- 153 ROMAN ROSZKO Russian language resources and tools on the Clarin-Pl website
- 177 TOMASZ SZUTKOWSKI In search for collocations and their translational equivalents with the use of electronic tools and corpus data
- 195 ALEKSANDRA ZIĘBA Parallel corpus InterCorp in the process of establishing Polish-Russian pairs of metaoperators

* * *

- 215 YEVHEN LIASHCHEVSKYI Interest in India and Hindu texts in Russian literature from the end of the 18th Century to the middle decades of the 20th Century
- 224 GRZEGORZ PRZEBINDA Three Meetings with Andrei Sinyavsky (Paris — Geneva — Kraków, 1987–1993)
- 249 ANDRZEJ ROMANOWSKI When the state breaks the law. . .
GRZEGORZ PRZEBINDA Interview with Mariya Rozanova and Andriey Siniavsky (november 1993)

REVIEWS

- 261 ALEKSANDRA SZYMAŃSKA Валентина Е. Головчинер, Татьяна Л. Веснина, *Комическое в пьесах М. Булгакова 1920-х годов*, Издательство Томского университета, Томск 2021
- 264 GRAŻYNA MAROSZCZUK *Gustaw Herling-Grudziński. Między Wschodem a Zachodem*, eds. Grzegorz Przebinda, Bartosz Gołąbek, Wojciech Gruchała, Wydawnictwo Humanistyczne Pigionium, Karpacka Państwowa Uczelnia w Krośnie, Krosno 2021
- 275 JELENA VAZHENINA RUSSIAN PHRASOLOGY AS FATE
ANATOLIY ZAHNITKO Анатолий Жуков, *Очерки по фразеологической семантике: монография*, 2-е издание, переработанное и дополненное, Издательство РУСАЙНС, Москва 2020
- 283 About the authors



ANDRZEJ CHARCIAREK

 ORCID: <https://orcid.org/0000-0001-5651-5730>

Uniwersytet Śląski w Katowicach

WSTĘP

Językoznawstwo jako dyscyplina nie zamknęło się w ujęciach strukturalistycznych czy formalistycznych i wciąż poszukuje nowych perspektyw badawczych. Początkowe skupienie uwagi na zagadnieniach systemowo-formalnych z czasem ustąpiło miejsca optyce poszerzonej o czynniki pozasystemowe. Stąd pojawienie się takich dziedzin w obrębie badań językoznawczych, jak socjolingwistyka, psycholingwistyka, pragmalingwistyka, etnolingwistyka czy mediolingwistyka. Wszystkie one wniosły i wciąż wnoszą wiele do opisu funkcjonowania języka w tekstach w określonym kontekście społeczno-kulturowym.

Nie inaczej jest z językoznawstwem korpusowym, które na dobre rozwinęło się w latach osiemdziesiątych ubiegłego wieku. Jego początki datują się od końca lat sześćdziesiątych i są związane z powstaniem pierwszego korpusu komputerowego zawierającego teksty w amerykańskiej odmianie angielszczyzny. Nie dziwi zatem fakt, że amerykańskie i angielskie środowiska naukowe miały największe zasługi dla rozwoju zarówno metodologii językoznawstwa korpusowego, jak i zasad tworzenia korpusów tekstów.

Początki narodowych korpusów języków słowiańskich to już lata dziewięćdziesiąte XX i pierwsza dekada XXI wieku. Wtedy to powstają narodowe korpusy języka czeskiego, polskiego czy rosyjskiego. Co trzeba podkreślić, rozpoczęcie prac nad ich budową nie było równoznaczne z ich udostępnieniem, do którego dochodziło zwykle po upływie wielu lat.

Dla przykładu, choć prace nad Narodowym Korpusem Języka Czeskiego (czes. *Český národní korpus*) rozpoczęły się w 1994 roku, to on sam (pierwszy korpus referencyjny) został udostępniony na początku trzeciego tysiąclecia.

Narodowy Korpus Języka Polskiego to wspólne przedsięwzięcie Instytutu Podstaw Informatyki PAN, Instytutu Języka Polskiego PAN, Wydawnictwa Naukowego PWN oraz Zakładu Językoznawstwa Komputerowego i Korpusowego Uniwersytetu Łódzkiego. I choć sam korpus został udostępniony w 2012 roku w rezultacie realizacji projektu w latach 2007–2011, to wspomniane instytucje gromadziły dane korpusowe i opracowywały narzędzia korpusowe już w latach dziewięćdziesiątych.

Publiczne udostępnienie do Narodowego Korpusu Języka Rosyjskiego (ros. Национальный корпус русского языка) nastąpiło w 2004 roku, a poprzedziły go prace prowadzone od 2001 roku przez wiele instytucji, których działania koordynował Instytut Języka Rosyjskiego im. W.W. Winogradowa Rosyjskiej Akademii Nauk w Moskwie.

Mimo że przywołałem tylko te trzy słowiańskie korpusy narodowe, nie zapominam o innych, których najpełniejszy chyba opis znajdziemy w monografii *Praktyczny przewodnik po korpusach języków słowiańskiej* (2014) pod redakcją Mileny Hebal-Jeziarskiej. Opracowanie to unaocznia, jak zróżnicowany jest zarówno stan badań korpusowych, jak i samych korpusów poszczególnych języków słowiańskich.

Uwzględniając okresy, w jakich kształtowały się inne dziedziny, na przykład pragmalingwistyka, można mówić o początkowym stadium rozwoju lingwistyki korpusowej. Jednak w odróżnieniu od poprzednich metodologii, które czerpały z osiągnięć innych dyscyplin (na przykład psycholingwistyka z psychologii, socjolingwistyka z socjologii), wnosi ona do prowadzonych analiz coś zupełnie nowego — metody ilościowe i empiryczne dotąd rzadko stosowane w badaniach lingwistycznych, a tak popularne w naukach ścisłych. W ten sposób lingwistyka korpusowa stanowi nowe podejście w metodologii badań — podejście statystyczne. Ponadto korzystanie z zasobów i narzędzi korpusowych umożliwia łączenie metod ilościowych i jakościowych.

Potwierdzenie wartości badań korpusowych znajdują Czytelnicy tego „korpusowego” numeru „Przeglądu Rusycystycznego” w poszczególnych jego artykułach, które obrazują prawdziwy wachlarz możliwości, jakie korpusy stwarzają badaczom języka. Wśród autorów znaleźli się językoznawcy zarówno tworzący korpusy językowe, jak i wykorzystujący je do celów badawczych. Zespala te wszystkie teksty obecność w każdym z nich języka rosyjskiego. Nie ma w tym przypadku, ponieważ jest to zgodne z profilem czasopisma adres-

WSTĘP

wanego głównie do rusycystów. To właśnie z myślą o nich zostali zaproszeni do jego współtworzenia Autorzy, którzy zechcieli podzielić się swoimi doświadczeniami z pracy z zasobami i narzędziami korpusowymi. Wszystkim im — z Lublina, Olsztyna, Opola, Szczecina, Tampere, Warszawy i Katowic — chciałbym gorąco podziękować za współpracę i łączące nas przekonanie, że środowisko rusycystyczne takiego opracowania potrzebuje. Mam tu na uwadze zwłaszcza młode pokolenie rusycystek i rusycystów — badaczy, którzy poszukują jeszcze swojej tożsamości naukowej i którzy może znajdą w korpusach nie tyle obiekt naukowej fascynacji, ile skuteczną metodę badania języka stanowiącą rozwiązanie alternatywne wobec tradycyjnych modeli teoretycznych i koncepcji metodologicznych.



EWA BIAŁEK

<https://orcid.org/0000-0003-3868-5001>

Uniwersytet Marii Curie-Skłodowskiej, Lublin

POLIFUNKCYJNOŚĆ KORPUSU JĘZYKOWEGO (NA PRZYKŁADZIE PLATFORMY SKETCH ENGINE I NARODOWEGO KORPUSU JĘZYKA ROSYJSKIEGO)

POLYFUNCTIONALITY OF LANGUAGE CORPUS

(ON THE MATERIAL OF THE SKETCH ENGINE AND THE RUSSIAN NATIONAL CORPUS)

The aim of the article is to show the possibilities of using the corpus in scientific research and foreign language teaching. The paper discusses the advantages of the Russian National Corpus and the corpus ruTenTen11 (Sketch Engine). The material presented in the corpus, which is a modern language tool, can be used for investigation of language changes as well as the linguistic worldview. The corpora record the history of words, document the changes in words' meanings. The author of the paper analyzes the lexeme and concept Демократия in Russian, the meanings of the new borrowing контент and the verb озвучить. In addition, the use of the language corpora in lexicography is analyzed, the author also compares the knowledge about the language and the culture that is contained in the corpora and the associative dictionary. The corpora provide useful information about semantic features of words and their collocations. By the patterns of collocability it is possible to reconstruct the linguistic worldview. One of the conclusions is that linguistic data extracted from the corpora can supplement traditional dictionaries.

Z HISTORII KORPUSÓW

Korpusy językowe, które zaczęto popularyzować od drugiej połowy XX w., w dwóch ostatnich dekadach zauważalnie upowszechniły się w badaniach lingwistycznych, leksykografii i (glotto)dydaktyce. Pierwsze korpusy obejmowały zasoby angielszczyzny. Pionierskim projektem dotyczącym angielszczyzny amerykańskiej jest Korpus Browna (ang. Brown University Standard Corpus of Present-Day American English), który udostępniono w latach 1963–1964¹. Po-

¹ BROWN Corpus, <https://www.sketchengine.eu/brown-corpus/> (11.10.21), zob. też E. Hajnicz, *Najbardziej znane korpusy tekstów. Opracowanie przeglądowe Nr 1021*, Instytut Podstaw Informatyki PAN, Warszawa 2011, s. 11–18.

wstanie korpusów słowiańskich przypada na przełom XX i XXI wieku. Prace nad Narodowym Korpusem Języka Rosyjskiego rozpoczęły się w 2001 roku, korpus udostępniono trzy lata później. Prace nad Narodowym Korpusem Języka Polskiego prowadzono od roku 2006². Dmitrij W. Siczynawa, pisząc o genezie korpusu rosyjskiego, wspomina o licznych seminariach, konferencjach i innych przedsięwzięciach z zakresu komputerowej obróbki i automatyzacji danych tekstowych poprzedzających utworzenie Narodowego Korpusu Języka Rosyjskiego³. Na dany korpus, poza najczęściej eksplorowanym zbiorem podstawowym (korpus podstawowy) obejmującym szeroki przekrój tekstów od XVIII do XXI wieku (proza, publicystyka, prasa, teksty naukowe, religijne, urzędowe, użytkowe), składa się korpus tekstów prasowych (uruchomiony w 2010 roku), poetyckich (2006), korpus równoległy (z dwutekstami), dialektologiczny (2005), syntaktyczny, dydaktyczny, języka mówionego (2007), akcentologiczny (2008), multimedialny (2010), a także korpus prasy regionalnej i zagranicznej (2015) oraz korpus historyczny⁴. Władimir A. Płungjan jako jedną z zalet korpusu postrzega różnorodność zasobów tekstowych ukazującą żywą i dynamiczną ruszczyznę⁵. Wykaz publikacji autorów rosyjskich zaangażowanych w prace nad korpusem liczy ponad 100 prac⁶. Kolekcje tekstów ujęte w korpusie oraz zawarte w nich dane stały się podstawą opracowań poświęconych zarówno konceptom językowo-kulturowym, leksyce rodzimej i zapożyczonej, zasobom frazeolo-

² O genezie korpusu zob. B. Lewandowska-Tomaszczyk, M. Bańko, R.L. Górski, M. Łaziński, P. Pęzik, A. Przepiórkowski, *Narodowy Korpus Języka Polskiego: geneza i dzień dzisiejszy*, w: A. Przepiórkowski, M. Bańko, R.L. Górski, B. Lewandowska-Tomaszczyk (red.), *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa 2012, s. 3–10.

³ Д. В. Сичинава, *Национальный корпус русского языка: очерк предыстории*, w: В. А. Плунгян (отв. ред.), *Национальный корпус русского языка: 2003–2005. Сборник статей*, Индрик, Москва 2005, s. 21–30, <https://ruscorpora.ru/new/sbornik2005/03sitch.pdf> (11.10.21). Wśród współtwórców korpusu są moskiewscy lingwiści z Instytutu Języka Rosyjskiego im. W. Winogradowa Rosyjskiej Akademii Nauk, Instytutu Językoznawstwa RAN oraz specjaliści z uczelni z całej Rosji (Sankt Petersburg, Kazań, Woroneż, Saratow), zob. uczestnicy projektu: <https://ruscorpora.ru/new/corpora-team.html> (11.10.21).

⁴ Dane ze strony: Narodowy Korpus Języka Rosyjskiego, <https://ruscorpora.ru/new/> (11.10.21).

⁵ В. А. Плунгян, *Зачем нужен Национальный корпус русского языка? Неформальное введение*, w: В. А. Плунгян (отв. ред.), *Национальный корпус русского языка: 2003–2005. Сборник статей*, Индрик, Москва 2005, s. 8–11 (6–20), <https://ruscorpora.ru/new/sbornik2005/02plu.pdf> (11.10.21).

⁶ Zob. publikacje: <https://ruscorpora.ru/new/corpora-biblio.html> (11.10.21).

gicznym i paremiologicznym, jak i zagadnieniom słowotwórczym, glottodydaktycznym, praktyce wydawniczej i in.⁷ Ta wielość tematów jest odpowiedzią na pytanie, czy dany korpus bądź jakikolwiek inny jest użyteczny i kto może z niego czerpać korzyści naukowe⁸. Objętość zbiorów językowych i reprezentatywność danych pozwalają, zdaniem Jacka Walińskiego, traktować korpusy jako podstawę „do wydawania sądów o języku”⁹. Niektórzy lingwiści idą dalej i korpusy językowe wiążą z nową ideologią badania języka (ideologia korpusowa)¹⁰.

Badacze ruszczyzny coraz częściej sięgają też do platformy korpusowej Sketch Engine¹¹ umożliwiającej przeszukiwanie gotowych korpusów w różnych językach europejskich i nieeuropejskich, jak również tworzenie korpusów własnych¹². Czym jest Sketch Engine? Odpowiedź zawiera ten oto krótki opis: „The web service includes, as well as the core software, a large number of corpora pre-loaded and ‘ready for use’, and tools for creating, installing and managing your own corpora”¹³. Pierwszymi użytkownikami platformy Sketch Engine byli leksykografowie i wydawnictwa. Dziś, dzięki rozwojowi lingwistyki korpusowej oraz upowszechnieniu się korpusów, beneficjentami i użytkownikami platformy stały się instytucje naukowe, uczelnie. Na platformie Sketch Engine są dostępne takie narzędzia, jak Word Sketch (Collocations and word combinations), Word Sketch Difference (Compare collocations of two words), Thesaurus (Synonyms and

⁷ Zob. publikacje dostępne w bibliotece elektronicznej „KiberLeninka”, <https://cyberleninka.ru/> (11.10.21).

⁸ Możliwości naukowego i dydaktycznego zastosowania korpusów językowych pokazano w licznych pracach, w tym m.in. w monografii: B. Lewandowska-Tomaszczyk (red.), *Podstawy językoznawstwa korpusowego*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2005.

⁹ J. Waliński, *Typologia korpusów oraz warsztat informatyczny lingwistyki korpusowej*, w: B. Lewandowska-Tomaszczyk (red.), *Podstawy językoznawstwa korpusowego*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2005, s. 40.

¹⁰ В. А. Плу́нган, *Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики*, «Русский язык в научном освещении» 2008, nr 2 (16), s. 7–20, <http://rjano.ruslang.ru/ru/archive/2008-2/7-20> (11.10.21).

¹¹ Sketch Engine access funded by ELEXIS: 2018–2022, <https://www.sketchengine.eu/elexis/> (11.10.21).

¹² Jest to około 90 języków, <https://www.sketchengine.eu/corpora-and-languages/> (11.10.21).

¹³ A. Kilgarriff, V. Baisa, J. Bušta i in. *The Sketch Engine: ten years on*, „Lexicography. Journal of ASIALEX” 2014, vol. 1, nr 1, s. 8 (7–36), <https://link.springer.com/article/10.1007/s40607-014-0009-9> (11.10.21). Adam Kilgarriff był współtwórcą platformy Sketch Engine.

similar words), Concordance (Examples of use in context), Keywords (Terminology Extraction) i in. Za ich pomocą w sposób automatyczny są generowane dane językowe stanowiące punkt wyjścia dla badacza. Wybrane dane można zwizualizować w postaci obrazów graficznych w formacie SVG, PDF, PNG (zob. dalej).

W artykule pokażę podstawowe sposoby wykorzystania korpusów językowych w analizie słownictwa, językowego obrazu świata, jak również dla celów praktyki leksykograficznej. Opracowanie to stanowi swego rodzaju zapis refleksji własnych pojawiających się podczas praktycznego użytkowania wskazanych narzędzi zarówno w pracy badawczej, jak i dydaktycznej. Zasoby Narodowego Korpusu Języka Rosyjskiego były przeze mnie wykorzystywane podczas opracowywania słownika *Dyplomacja i polityka. Rosyjsko-polski słownik przekładowy*, А–Д¹⁴ oraz publikacji okołosłownikowych o charakterze metaleksykograficznym, jak również w pracy nauczyciela akademickiego¹⁵.

KORPUSY A JĘZYKOWY OBRAZ ŚWIATA

Użyteczność korpusów językowych zilustruję poprzez analizę leksemu демократия z wykorzystaniem platformy Sketch Engine¹⁶. Korpusy, podobnie jak słowniki frekwencyjne, znajdują zastosowanie w badaniu językowego obrazu świata¹⁷. Odwołania do danych korpusowych stanowią stały element kompozycji publikacji z *Leksykonu aksjologicznego Słowian i ich sąsiadów* (t. I–V). Pozyskana w ten sposób wiedza językowa stanowi podstawę definicji kognitywnej i profilowania¹⁸. Demokracja, wraz z godnością człowieka, wolnością, równo-

¹⁴ E. Białek, *Dyplomacja i polityka. Rosyjsko-polski słownik przekładowy*, А–Д, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 2019.

¹⁵ Do narzędzi korpusowych odwołuję się podczas różnych typów zajęć: Terminologia i terminografia rosyjska, Tendencje rozwojowe współczesnego języka rosyjskiego, Analiza przekładu, Gramatyka opisowa współczesnego języka rosyjskiego i in.

¹⁶ Platformę Sketch Engine wykorzystuję także w badaniu conceptów ТЕРПИМОСТЬ, ТОЛЕРАНТНОСТЬ w języku rosyjskim (badania prowadzone w 2021 roku).

¹⁷ Zob. A. Pawłowski, *Metodologiczne podstawy wykorzystania słowników frekwencyjnych w badaniu językowego obrazu świata*, w: A. Pajdzińska, P. Krzyżanowski (red.), *Przestrzeń w językowym obrazie świata*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 1999, s. 81–99.

¹⁸ Zob. J. Bartmiński, *Czym różni się Leksykon aksjologiczny od słownika językowego i encyklopedii?*, w: S. Niebrzegowska-Bartmińska, J. Szadura, B. Żywicka (red.), *Wartości w językowo-kulturowym obrazie świata Słowian i ich sąsiadów*, 4, *Słownik językowy — leksykon — encyklopedia w programie badań*

ścią, praworządnością i prawami człowieka, pojawia się wśród sztandarowych wartości europejskich (UE)¹⁹. Pojęcie demokracji ma też znaczenie dla społeczeństwa rosyjskiego, które w latach pierestrojki weszło na drogę przemian. Pojęcie to jest różnie konceptualizowane i wartościowane (od ustroju pożądanego do ustroju postrzeganego jako nieład bądź system ideologicznie i mentalnościowo obcy)²⁰. Zasoby korpusu w języku rosyjskim Russian Web 2011 (ruTenTen11) obejmujące 14 553 856 113 słów pokazują funkcjonowanie tego wyrazu w tekstach, a zarazem umożliwiają prześledzenie sposobu postrzegania demokracji jako konceptu z różnych perspektyw: po pierwsze, demokracja w ujęciu historycznym (od najwcześniejszej demokracji antycznej do demokracji europejskiej XX i XXI wieku), po wtóre, demokracja na świecie (przemiany), po trzecie, kształtowanie się demokracji w Rosji (por. народная демократия — demokracja ludowa w ZSRR²¹, управляемая демократия ‘demokracja sterowana’)²², po czwarte, społeczeństwo rosyjskie o demokracji „cudzej” i „własnej”. Do analizy łączliwości wyrazu демократия i zrekonstruowania na tej podstawie konceptu demokracji²³ zastosowano narzędzie Word Sketch służące do ekstrakcji kolokacji i innych analizyzmów. Wygenerowane wyniki są przyporządkowane 46 modyfikatorom. Wybrane z nich wraz z formacjami wyrazowymi przedstawiam poniżej:

porównawczych, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 2018, s. 9–23.

¹⁹ Zob. Cele i wartości UE, https://europa.eu/european-union/about-eu/eu-in-brief_pl (11.10.21).

²⁰ Koncept demokracji w języku i kulturze rosyjskiej szerzej omawiam w artykule: E. Białek, A. Yudin, *Koncept ДЕМОКРАТИЯ в языке rosyjskim: dane S–A–T*, w: J. Bartmiński, M. Grzeszczak, A. Rudenka (red.), *Leksykon aksjologiczny Słowian i ich sąsiadów*. Tom: *Demokracja*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 2022 (w druku).

²¹ *Demokracja ludowa* — „polit. forma rządów realizowana w ZSRR i innych krajach komunistycznych, opierająca się na faktycznej władzy jednej partii komunistycznej przy zachowaniu pozorów formalnych form demokracji parlamentarnej”. P. Żmigrodzki (red.), *Wielki słownik języka polskiego PAN*, <https://wsjp.pl/haslo/podglad/31267/demokracja-ludowa/3021049/ustroj> (11.10.21).

²² Demokrację sterowaną dziennikarze objaśniają prostym językiem: „управляемая демократия — это когда голосует народ, а люди, находящиеся у власти, чуть-чуть выбор народа корректируют. В чью пользу? В свою, разумеется” (В. Третьяков, *Диагноз: управляемая демократия*, «Независимая газета», 13.01.2000, https://www.ng.ru/editor/2000-01-13/1_diagnos.html (11.10.21).

²³ Przedstawiana analiza ma charakter przekrojowy i ogólny zarazem. Teksty zebrane w korpusie, jak wspomniano, poruszają wątek demokracji w różnych krajach i w różnych okresach.

subject_of: восторжествовала, расцветает, царит демократия; демократия кончается;

a_modifier: афинская, западноевропейская; образцовая, подлинная демократия; буржуазная, митинговая, народная, революционная, социалистическая, суверенная, тоталитарная, управляемая, электронная²⁴; внутрипартийная, рыночная демократия²⁵;

gen_modifies: торжество демократии; идеалы, колыбель, ростки, светоч демократии;

и/или: гласность и демократия, демократия и демократизация, демократия и вседозволенность, свободомыслие и демократия;

pp_v: демократия в Афинах, РФ, США, ЕС, Европе, Польше; демократия в кавычках;

object4_of: идеализировать, прославлять, сеять, строить демократию; дискредитировать, отвергать, растоптать демократию; навязывать, экспортировать демократию;

gen_modifier: демократия по-ельцински, по-путински, по-русски, западного образца;

pp_obj_k: на пути к демократии, шаг к демократии; апеллировать к демократии;

pp_obj_za: выступать, сражаться за демократию; борец за демократию;

pp_obj_o: болтовня, миф, сказки, споры о демократии; разглагольствовать о демократии;

pp_obj_v: верить в демократию, играть в демократию; сомневаться в демократии;

object2_of: учиться демократии; вредить, мешать, угрожать демократии;

pp_obj_c: заигрывание, эксперимент с демократией; воевать с демократией;

pp_na: демократия на Западе, на всей планете, на постсоветском пространстве;

pp_obj_na: атака на демократию, тест на демократию; посягать на демократию;

inst_modifies: что-л. зовётся, именуется, называется демократией;

pp_obj_ot: отказ от демократии; отойти, удалиться, устать от демократии;

pp_obj_do: дорасти, dozреть, созреть до демократии;

pp_obj_nad: глумление, насмешка над демократией; издеваться над демократией;

pp_dlya: демократия для богатых, бедных, трудящихся²⁶.

²⁴ Demokracja może oznaczać 'ustrój', np. подлинная демократия, lub 'państwo', np. молодая демократия. W zależności od kontekstu dane słowo wraz z tym samym określeniem może też odnosić się zarówno do 'ustroju', jak i 'państwa', por. образцовая демократия. Zob. znaczenia leksemu: В. В. Морковкин, Г. Ф. Богачёва, Н. М. Луцкая, *Большой универсальный словарь русского языка*, под ред. В. В. Морковкина, «АСТ-ПРЕСС КНИГА», «Словари XXI века», Москва 2016, s. 241.

²⁵ Demokracja w tych połączeniach ma znaczenie systemu panującego w jakiejś organizacji bądź obszarze (rynek).

²⁶ Przedstawiono wybrane modyfikatory i wygenerowane za ich pomocą wyniki.

Rozbudowane schematy kolokacyjne, poza użyteczną rejestracją łączliwościowych możliwości słowa, dostarczają wiedzy pozajęzykowej, tworzą też wielodzielny obraz demokracji. Zebrane przykłady pozwalają widzieć demokrację zarówno jako wartość/ustrój pozytywny, jak też jako wartość/ustrój niekorzystny oraz jako wartość/ustrój obcy (демократия западного образца), a także ustrój, system karykaturalny, ułomny (дефектная демократия). O demokrację jako wartość pozytywną toczy się walka, demokracja pojmowana w kategoriach dobra i ideału ostatecznie triumfuje nad odmienną ideologią i odmiennym systemem społecznym (umowna walka dobrego ze złem, por. демократия победила). Choć demokracja opiera się na zasadach, filarach (устои демократии), jej trwałość może być podana w wątpliwość (por. закончилась демократия, демократия изжила себя). Demokracja jest też konfliktogenna. Jej wdrażanie oraz sposoby rozumienia prowadzą do sporów, podważania zalet przez jej krytyków. Jest idealizowana przez jedną grupę (поборники, приверженцы демократии), a dyskredytowana przez drugą (враги демократии). Postrzeganie demokracji jako towaru eksportowego (por. экспортировать демократию, экспорт демократии) oznacza jej uprzedmiotowienie i materializację, a zarazem pejoratywny odbiór: demokracja jako system obcy, wiążący się z inną sferą wartości i kulturą, a nawet narzucany przez określone państwa w celu zmiany strefy wpływów i zawojowania nowych rynków (por. навязывать демократию).

Wyniki pozyskane za pomocą modyfikatora gen_modifies są przydatne do określenia tego, co stanowi oparcie demokracji lub jest jej wyrazem. Bastionami demokracji są określone instytucje (np. Верховный суд), kraje lub regiony (np. США, Великобритания), media (пресса, Интернет), a nawet sztuka, która pozwala na różne formy artystycznego wyrazu. Za cytadelę współczesnej demokracji uznaje się USA (цитадель мировой демократии США; wartościowanie tak pozytywne, jak i pejoratywne)²⁷, budynki Parlamentu Europejskiego (цитадель европейской демократии, por. niżej). Mianem cytadeli demokracji miejskiej ochrzczone Pałac Maryjski w Sankt Petersburgu (цитадель городской демократии)²⁸.

В цитадель европейской демократии можно сходить на экскурсию и даже посетить пленарное заседание [...]. (Sketch Engine)²⁹.

²⁷ Od wyników w postaci połączeń można przejść do konkordancji: fragmentów tekstów, w których przykłady zostały użyte, co pozwala na ich głębszą analizę.

²⁸ Jest to siedziba organu ustawodawczego miasta.

²⁹ Numer dokumentu (Document number) — 5438961.

Ustrój demokratyczny wymaga wzorców, w związku z czym poszczególne kraje z rozwiniętą demokracją są uznawane bądź same pozycjonują się jako wzorzec demokracji. Wspólnoty międzynarodowe oceniają kraje pod względem demokracji, w oficjalnych oświadczeniach wytykają braki i deficyt demokracji (дефицит демократии). Zamykając tę część artykułu, zauważę, że pełniejsza rekonstrukcja dowolnego konceptu na materiale konkretnego języka wymaga szczegółowszych badań, uwzględniających dane frekwencyjne. Dane te stanowią bowiem „językowe odbicie fragmentu wiedzy o świecie”³⁰.

KORPUSY A DAWNE W JĘZYKU

W oparciu o zasoby tekstowe z korpusów można prześledzić funkcjonowanie słownictwa w dłuższych odcinkach czasowych oraz stopniowo zachodzące w nim zmiany semantyczne. Dane te są komplementarne wobec danych słownikowych. Wyraz *бахилы*, dziś znany zwłaszcza w znaczeniu obuwia ochronnego (защитные, медицинские бахилы) lub roboczego (бахилы для рыбалки) przed ponad dwoma stuleciami oznaczał obuwie chłopskie³¹. Najstarsze poświadczenie w Narodowym Korpusie Języka Rosyjskiego (korpus podstawowy) dostarcza informacji o wyglądzie tego obuwia i materiale (przykład pierwszy), w zapisie wypowiedzi nosiciela dialektu/gwary (przykład drugi) zawarta jest z kolei informacja o funkcjonalności obuwia (korpus dialektologiczny):

Лаптей ни Башкирки, ни Башкирцы никогда не употребляют; но носят въ Уфимскомъ уѣздѣ упоки собственного кожевничества, а въ Исетской провинціи **бахилы**, у которыхъ только кайма у головъ кожаная, а прочее все суконное [И.И. Лепехин, *Продолженіе Дневныхъ записокъ путешествія академика и медицины доктора Ивана Лепехина по разнымъ провинціямъ Россійскаго государства въ 1770 году* (1770)].

// **бахилы** мужыкы — те вотлѣс — от ѳсеню да зимой — то ходятѣшо йѣтта — ка/ до колѣн // йешо выше сѣ кожаной сошѣут // и вот и [*Сарафаны носили, дубасы* (д. Мыс, Косинский район, Коми-Пермяцкий округ, Пермский край, 2003–2013)].

³⁰ A. Pawłowski, *Metodologiczne podstawy wykorzystania słowników frekwencyjnych...*, s. 90.

³¹ Zob. *Словарь Академии Россійской*, Часть 1, отъ А до Г. Императорская Академия наук, СПб. 1789, http://etymolog.ruslang.ru/doc/SAR1_A-G.pdf (11.10.21).

Dane tekstowe z XX wieku wskazują na rozbudowę znaczenia słowa (por. одноразовые бахилы, бахилы музейного типа — obuwie o funkcji ochronnej). W dialektach słowo to może oznaczać także obuwie wykonane z brezentu, nakładane na przykład na walonki³². Dane Sketch Engine uzupełniają wiedzę o typie obuwia i sferze jego zastosowania: войлочные, кожаные, матерчатые, меховые, резиновые бахилы (material); больничные, музейные, рыбацкие oraz велосипедные бахилы (nowy typ obuwia — ochraniacze kolarskie, nakładki na buty). Dany przykład zdaje się potwierdzać myśl Władimira A. Płungjana sformułowaną na początku ery korpusowej w Rosji (2008) o tym, że korpusy umożliwiają dostrzeganie subtelnie zachodzących zmian w semantyce i łączliwości wyrazów³³.

KORPUSY A NOWE W JĘZYKU

Wysoką użyteczność korpusów dostrzegam również w przypadku analizy słownictwa nowego.

Zapożyczenie контент, inaczej ‘treść, informacja’ stanowi obcojęzyczny synonim wyrazów содержание, информация. Zakres tej synonimicznej relacji określają definicje słownikowe³⁴. Wyraz został przejęty wraz z innymi terminami komputerowymi w dobie technologicznego boomu przypadającego na przełom XX i XXI w. (ang. *content*). We współczesnej ruszczyźnie dominuje ‘informatyczne’ znaczenie: treść strony, materiały zamieszczone na stronie (‘информационное наполнение сайта’), słowo może też oznaczać zawartość filmów, nagrań, tekstów (‘все размещаемые в Интернете материалы — текст, графические изображения, музыкальные и видеофайлы и т.п.’)³⁵. Definicja ze słownika ogólnego wskazuje na równie szeroki zakres zastosowania słowa: ‘Книжн. Содержание книги, статьи, пьесы. Анализировать к. сайта’³⁶. Pierwsze teksty

³² Zob. А. С. Герд (ред.), *Селигер. Материалы по русской диалектологии. Словарь. Выпуск 1, А–Г*, Издательство С.-Петербургского университета, СПб. 2003, s. 31.

³³ В. А. Плунгян, *Корпус как инструмент и как идеология:...*, s. 14. O zastosowaniu korpusów językowych zob. także: В. А. Плунгян, *Зачем нужен Национальный корпус русского языка?...*, s. 18.

³⁴ Zob. pr.: Е. Ю. Ваулина, *Информатика. Толковый словарь*, «Издательство «Эксмо», Москва 2005, s. 218.

³⁵ Tamże.

³⁶ С. А. Кузнецов (глав. ред.), *Большой толковый словарь русского языка*, Норинт, СПб. 2014 (1998), <http://www.gramota.ru/slovari/info/bts/> (11.10.21).

z Narodowego Korpusu Języka Rosyjskiego dostarczają użytkownikom wiedzy o przybliżonym czasie wejścia słowa wraz ze znaczeniem do językowego obiegu — najwcześniejsze konteksty pochodzą z 2000 roku (korpus podstawowy) i 2004 roku (prasowy). Nie należy mylić tej pożyczki z pożyczką wcześniejszą, której świadectwo zawiera ten oto przykład z danego korpusu:

Униженной мой поклон посылая, всепокорно доношу вам, что весма **контент** от получения сего числа милостиваго Вашего письма [С.Ф. Платонов, *Письма* (1910)].

Korpus multimedialny oraz akcentologiczny dostarczają wiedzy o akcencie i wymowie: контѐнт [тэ]³⁷. Analiza konkordancji doprowadzi badacza do wniosków na temat kolokacji, a od kolokacji do dalszych spostrzeżeń o znaczeniu. A zatem dany wyraz może oznaczać: 1) treść mającą konkretnego adresata (kryterium np. wiekowe) — взрослый контент; 2) treść wyspecjalizowaną (tematyka, cel) — развлекательный, рекламный, учебный контент; 3) treść o określonym formacie, zamieszczoną w pewnej przestrzeni, na nośniku: мобильный, цифровой, электронный контент, por. także контент канала, портала, сайта, телеканала; 4) treść podlegająca ocenie: некачественный; 5) treść pozyskiwaną legalnie/nielegalnie: пиратский контент; 6) treść podlegającą przenoszeniu, przekazywaniu: ретвитить контент.

Z listy wyników na platformie Sketch Engine wygenerowanych za pomocą narzędzia Word Sketch (modyfikator: a_modifier) wyekscerpowano typowe kolokaty przymiotnikowe wspomagające identyfikację sfery użycia, jak również wartościujące dane słowo-pojęcie: 1) ocena negatywna (treść pozyskiwana nielegalnie; treść sprzeczna z prawem, por. wyżej) — ворованный контент; запрещённый, противозаконный, противоправный контент; 2) ocena pozytywna (treść o wysokiej jakości; treść dostępna; treść wpływająca na stan emocjonalny odbiorcy) — высококачественный, эксклюзивный контент; незапрещённый контент; позитивный контент. Z zastosowaniem modyfikatora gen_modifier wygenerowane zostały kolokaty w formie dopełniacza, m.in.: блога, журнала, интернет-магазина, патча, рассылки, страницы, Рунета. Słowo można od-

³⁷ W słownikach wskazywany jest też wcześniejszy, wyparty już wariant akcentuacyjny: контѐнт (akcent na pierwszej sylabie), zob. Е.Ю. Ваулина, *Информатика. Толковый словарь*,...

nieść i do wytworów ludzkiej działalności w formie tradycyjnej, np.: контент бумажного журнала. W zasobach ogólnych Narodowego Korpusu Języka Rosyjskiego odnotowano wystąpienia badanego słowa także w literaturze pięknej. Zapożyczenie pojawia się w utworze Wiktora Pielewina: „А вот контент-сомелье из тебя выйдет [Виктор Пелевин, *S.N.U.F.F.* (2011)]. Dane obu korpusów wzajemnie się uzupełniają.

W analizie konceptów, poza narzędziami lingwistyki korpusowej, są stale stosowane tradycyjne metody eksperymentalne (ankiety). Marina W. Klimowa, Uljana I. Turko i Galina N. Abreimowa poddały leksem контент badaniu eksperymentalnemu. Na podstawie ankiet sformułowały własną definicję słowa-konceptu, która uwzględnia takie komponenty, jak: lokalizacja treści, forma jej przekazu (słowo, obraz), ocena jakości, ważność dla internauty (por. przedstawiony wyżej opis znaczenia wyrazu oparty na łączliwości wyrazowej, dane korpusowe):

это размещенная в Интернете интересная информация (возможно, неподтвержденная), передающаяся разными способами (печатный текст, иллюстрация, аудио, видео, их сочетание), которая рассчитана на массового читателя и на ознакомление с которой современная молодежь тратит 1/3 активного времени в течение суток³⁸.

Platforma Sketch Engine porządkuje wiedzę o potencjale semantycznym czasownika озвучить (znaczenie podstawowe — проф., озвучить фильм, киноленту; озвучить актёра)³⁹, który w nowym znaczeniu dołączył do listy czasowników mowy oraz dotyczących wymiany informacji (por. заявить, известить, огласить, сказать, сообщить, уведомить; высказаться; довести до сведения, etc.). To rozszerzenie semantyki można zobrazować za pomocą listy rzeczowników wygenerowanych na platformie Sketch Engine (korpus

³⁸ М. В. Климова, У. И. Турко, Г. Н. Абреимова, *О значении слова контент в языковом сознании молодежи*, «Научный диалог» 2021, № 7, s. 91–107, <https://cyberleninka.ru/article/n/o-znachenii-slova-kontent-v-yazykovom-soznanii-molodezhi> (21.10.21). Autorki, poza podstawowym pytaniem *Что такое контент?*, postawiły grupie 50 studentów także i takie pytania: *Каковы основные функции контента? Какая форма представления контента для Вас наиболее интересна?* (Tamże, s. 96).

³⁹ Zob. С. А. Кузнецов (глав. ред.), *Большой толковый словарь русского языка...*; Н. Ю. Шведова (отв. ред.), *Толковый словарь русского языка с включением сведений о происхождении слов*, Издательский центр «Азбуковник», Москва 2008, s. 559–560.

ruTenTen11) za pomocą modyfikatora object4 (rzeczownik jako baza imienna kolokacji): озвучить вердикт, идею, инициативу, итоги, кандидатуру, мысль, мнение, намерение, недовольство, обеспокоенность, опасения, повестку дня, позицию, послание, предложение, просьбу, статистику, тезис, ультиматум, ценник. Rzeczowniki są zróżnicowane — od nazw uczuć i opinii, aktów mowy do nazw dokumentów. Modyfikator subject typuje podmiot. Nazwy profesji i piastowanych urzędów wskazują zarazem na status społecznego nadawcy: глава, градоначальник, губернатор, министр, мэр, президент, премьер-министр, гендиректор, пр.:

Премьер **озвучил официальную точку зрения** правительства [...]. (Sketch Engine)⁴⁰.

Ujednolicenie pod względem semantycznym zestawu nominacji wiąże się z tym, że czasownik wchodzi w skład żargonu zawodowego polityków i urzędników. To przekazywanie informacji może odbywać się zarówno w formie ustnej, jak i pisemnej, ponadto nie tylko w oficjalnych okolicznościach (озвучить на брифинге, конференции, митинге, саммите, совещании, форуме), ale także w przestrzeni internetowej (озвучить что-л. в своём блоге). Na podstawie analizy tekstów spoza korpusu można dodać, że subiekt nie jest ograniczony jedynie do nominacji osób piastujących ważne stanowiska (administracje miast, państwa), por. Tik Tok-блогер, дизайнер озвучил что-л. (в эфире, соцсетях). Jednoznacznie jest to słowo modne i jak to słowo modne — nieco nadużywane. Choć krytycy użycia czasownika w tym znaczeniu widzą zagrożenie dla tradycyjnych czasowników mowy⁴¹, to częstość jego użycia przekłada się w konsekwencji na oswojenie nowej łączliwości oraz coraz większą tolerancję dla nowego kontekstu występowania (por. актёр озвучил фильм — министр озвучил проект — докладчик озвучил результаты опроса — писатель озвучил мысль — блогер озвучил секрет успеха — певец озвучил свои планы).

KORPUSY A LEKSYKOGRAFIA

Pozyskane z korpusów dane można przenieść do słownika, na przykład budując artykuł hasłowy идентичность. Porównam dane zebra-

⁴⁰ Numer dokumentu (Document number) — 17354753.

⁴¹ Zob. О. Северская, *Говорим по-русски с Ольгой Северской*, Слово / Slovo, Москва 2004, s. 46–47.

ne przeze mnie samodzielnie na podstawie materiałów dostępnych na stronie MSZ Rosji⁴² oraz dane, które badacz mógłby pozyskać z platformy korpusowej Sketch Engine (Word Sketch, Basic — wersja podstawowa, gotowe kolekcje tekstów)⁴³. Oto samodzielnie utworzona lista kolokacji wyrazu *идентичность* (wybrane), w których realizowane jest znaczenie tożsamości jako poczucia wspólnoty, jedności (z narodem, jego przeszłością i wyznawanymi wartościami). Dane znaczenie wpisuje się w dyskurs dyplomatyczny, polityczny, jest w nim dominujące: *историческая, культурная, собственная национальная, национально-культурная, цивилизационная, этническая, этнокультурная, языковая идентичность*⁴⁴. Korpus ruTenTen11 zawiera, choć częściowo zbieżną z przedstawioną, to jednak bogatszą listę, m.in. *гендерная, групповая, корпоративная, личностная, маскулинная, наднациональная, надэтническая, национально-культурная, половая, профессиональная, общегражданская, общероссийская, цивилизационная, этническая*. Szersza łączliwość bazy, ośrodka kolokacji (rzeczownik) wynika z aktywizacji w tekstach znaczenia ‘tożsamość własna, świadomość siebie’, teksty te współtworzą inne typy dyskursów. Leksem *идентичность* we wskazanych znaczeniach w leksykografii rosyjskiej jest praktycznie nieobecny⁴⁵, a zatem dane tekstowe i korpusowe można przekuć na artykuł hasłowy w słowniku jedno- lub dwujęzycznym odpowiadający faktycznemu znaczeniu słowa i wiedzy o tym koncepcie we współczesnej ruszczyźnie.

KORPUSY A SŁOWNIKI ASOCJACYJNE

Wyniki z korpusów można zestawić z zawartością słowników asocjacyjnych. Niekiedy to zestawienie wypada na korzyść mniejszego objętościowo zbioru, jakim jest słownik. Listy słów-bodźców uwzględniają bowiem prymarne skojarzenia słowne obudowane kulturowo — w trakcie badania eksperymentalnego są one podawane jako pierwsze. Skojarzenia te tworzą model językowego obrazu świata nosiciela

⁴² Ministerstwo Spraw Zagranicznych Rosji, <https://www.mid.ru/ru/home> (11.10.21).

⁴³ Zestawienie ma charakter ilustracyjny, z platformy Sketch Engine w ramach projektu ELEXIS można korzystać w celach niekomercyjnych.

⁴⁴ Przykłady wejdą w skład kolejnej części słownika. E. Białek, *Dyplomacja i polityka...*

⁴⁵ Zob. hasło *идентичный* w słowniku: С.А. Кузнецов (глав. ред.), *Большой толковый словарь русского языка...*

danego języka i przedstawiciela danej kultury⁴⁶. Na liście słów-reakcji zarejestrowanych w słowniku YEVRAS⁴⁷ dla słowa-bodźca медведи odnotowano wyrazy określające jego cechy fizyczne (белые, бурые, большие, сильные), gatunek (гризли, панда, полярные), miejsce zamieszkiwania (в лесу, в бору, берлога, север, Камчатка, Дальний Восток, Сибирь), typowe zachowanie (спят, ловят рыбу), pożywienie (мёд), los zniewolonych osobników (зоопарк, клетка). Pojawia się również skojarzenie z lat dzieciństwa (плюшевые — плюшове миsie). Szczególnie cenne są reakcje odnoszące się do kultury rosyjskiej: Шишкин, косолапый, картина, конфета/конфеты, шоколадки. Niedźwiedzie pojawiają się na obrazie *Poranek w sosnowym lesie* Iwana Szyszkiina (1889), ten motyw widnieje na opakowaniu produktu znanych zakładów cukierniczych Krasnyj Oktiabr' (kultowe cukierki Мишка Косолапый)⁴⁸. Przymiotnik косолапый to stały epitet niedźwiedzia w folklorze, literaturze — w zasobach tekstowych Sketch Engine towarzyszy on rzeczownikom мишка, медвеженок, медведь, зверь. W eksperymencie lingwistycznym pada ponadto nazwisko prezydenta Dmitrija Miedwiediewa (urobione od nazwy zwierzęcia), jak również nazwa partii Jedna Rosja (Единая Россия) z postacią niedźwiedzia na logotypie oraz nazwa bijącego rekordu oglądalności filmu animowanego *Masza i Niedźwiedź*. W korpusie ruTenTen11 lista towarzyszących wyrazów jest znacznie szersza, jednakże rozproszenie danych utrudnia tak szybkie i kompaktowe zrekonstruowanie obrazu niedźwiedzia jako gatunku czy w aspekcie kulturowym. Nie przeczy to oczywiście zaletom edukacyjnym korpusu językowego, dzięki któremu ten obraz można znacznie pogłębić.

KORPUSY A SYNONIMY I PARONIMY

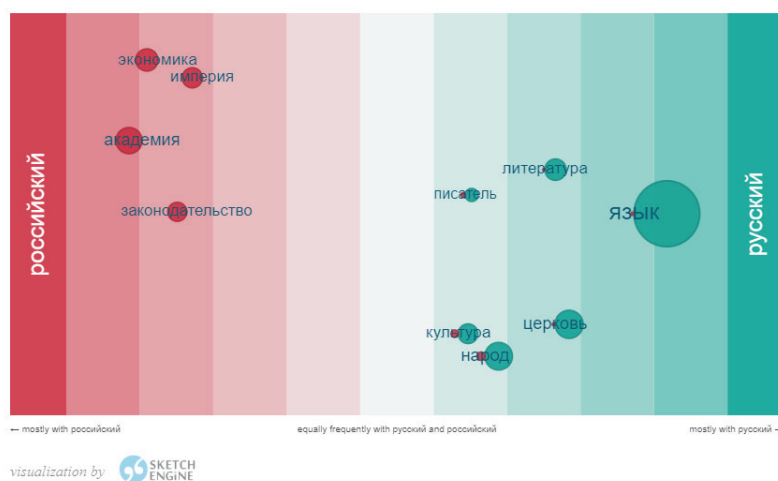
Narzędzie Word Sketch Difference jest użyteczne w przypadku porównywania synonimów i paronimów. Podstawą analizy wybranych

⁴⁶ Н. В. Уфимцева, *Языковая картина мира как предмет психолингвистического исследования*, w: D. Szumska, K. Ozga (red.), *Язык и метода*, 2, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2015 (data publikacji online: 2016), s. 249–256, <https://www.ejournals.eu/Language-and-Method/2015/2015/art/6683> (21.10.21).

⁴⁷ Г. А. Черкасова, Н. В. Уфимцева, *Русский региональный ассоциативный словарь-тезаурус ЕВРАС, Том 1, От стимула к реакции*, Москва 2014, <https://iling-ran.ru/web/ru/publications/evras> (11.10.21).

⁴⁸ Zob. też: „Красный Октябрь»: путь от кондитерской фабрики до арт-кластера, 30.09.2011, <https://reality.ria.ru/20110930/396734786.html> (11.10.21).

przez użytkownika wyrazów są listy ich kolokacji (dla przykładu, można dokonać porównania przymiotników: русский — российский, демократический — демократичный, дипломатический — дипломатичный)⁴⁹. Dostępna jest również opcja wizualizacji danych (poniżej). Podobieństwa i różnice w wynikach są oznaczone kolorami. Relacje te są wielorakie: przewaga bądź nieobecność jednego z wyrazów w określonych połączeniach, równoległe współtworzenie tych samych kolokacji. Zalety wizualizacji są oczywiste — mimo uproszczonego obrazu do użytkownika korpusu dociera czytelna elementarna informacja, do wykorzystania w pracy dydaktycznej, na przykład: русская культура — российская экономика, демократическая оппозиция — демократичные джинсы; демократичная атмосфера.



KONKLUZJA

Korpusy językowe stanowią wielofunkcyjne narzędzie stosowane w badaniach nad językiem i tekstem: w ujęciu diachronicznym i synchronicznym, w odrębnych gatunkach i dyskursach, w odmianie pisemnej i ustnej. Skala wykorzystania korpusu oraz ocena użyteczności zgromadzonych w nim zasobów tekstów pozostaje w bezpośrednim

⁴⁹ Porównanie ostatniej pary przymiotników przeprowadziłam uprzednio na materiale danych słownikowych oraz korpusowych, zob. Э. Бялэк, *К вопросу об узусе прилагательных «дипломатичный» и «дипломатический»*, Международный научный журнал «Филологические науки» 2014, № 4, s. 37–47.

związku z założonym celem funkcjonalizacji danych. Zawarte w artykule rozważania bazują na praktycznym użyciu podstawowych narzędzi korpusowych. Omówione przykłady pokazują polifunkcyjny charakter korpusów, które sprawdzają się nie tylko w automatycznym pozyskiwaniu list kolokacji, ale też wspierają analizę konceptów i rekonstrukcję językowego obrazu świata. Korpusowy kierunek badań przeżywa dziś rozkwit, jednak nie powinien wyprzeć całkowicie wcześniejszych metod czy ograniczać swobody badacza w indywidualnym – niekorpusowym – podejściu do słowa oraz tekstu.

REFERENCES

- Bartmiński, Jerzy. “Czym różni się Leksykon aksjologiczny od słownika językowego i encyklopedii?”. *Wartości w językowo-kulturowym obrazie świata Słowian i ich sąsiadów*, 4, *Słownik językowy – leksykon – encyklopedia w programie badań porównawczych*. Eds. Niebrzegowska-Bartmińska, Stanisława, and Szadura, Joanna, and Żywicka, Beata. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 2018. 9–23.
- Białek, Ewa. *Dyplomacja i polityka. Rosyjsko-polski słownik przekładowy*. A–D. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 2019.
- Białek, Ewa. “К вопросу об узусе прилагательных «дипломатичный» и «дипломатический».” *Международный научный журнал «Филологические науки»*, 2014, no. 4: 37–47, <https://filolnauki.ru/ru/archive/904/3281> [Бялэк, Эва. “К вопросу об узусе прилагательных «дипломатичный» и «дипломатический».” *Международный научный журнал «Филологические науки»* 2014, no. 4: 37–47, <https://filolnauki.ru/ru/archive/904/3281> (11.10.21)].
- Białek, Ewa, Yudin, Aleksey. “Koncept ДЕМОКРАТИЈА в языке росыјским: дане S–A–T.” *Leksykon aksjologiczny Słowian i ich sąsiadów*. T.: *Demokracja*. Eds. Bartmiński, Jerzy, and Grzeszczak, Monika, and Rudenka, Alena. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 2022 (forthcoming).
- BROWN Corpus, <https://www.sketchengine.eu/brown-corpus/> (11.10.21).
- Cele i wartości UE, https://europa.eu/european-union/about-eu/eu-in-brief_pl (11.10.21).
- Cherkasova, Galina A., and Ufimtseva, Natal’ya V., *Russkiy regional’nyy assotsiativnyy slovar’-tezaurus YEVRAS*. T. 1: *Ot stimula k reaktsii*. Moskva, 2014, <https://iling-ran.ru/web/ru/publications/evras> [Черкасова, Галина А., and Уфимцева, Наталья В. *Русский региональный ассоциативный словарь-тезаурус ЕВРАС*. Том 1: *От стимула к реакции*. Москва, 2014, <https://iling-ran.ru/web/ru/publications/evras> (11.10.21)].
- Elektroniczna Biblioteka Naukowa „KiberLeninka”, <https://cyberleninka.ru/> (11.10.21).
- Gerd, Aleksandr S. (Ed.). *Seliger. Materialy po russkoy dialektologii. Slovar’. Vypusk 1*, A–G. Sankt-Peterburg: Izdatel’stvo S.-Peterburgskogo Universiteta, 2003 [Герд, Александр С. (Ed.). *Селигер. Материалы по русской диалектологии. Словарь. Выпуск 1*, A–Г. СПб.: Издательство С.-Петербургского университета, 2003].
- Hajnicz, Elżbieta. *Najbardziej znane korpusy tekstów. Opracowanie przeglądowe Nr 1021*. Warszawa: Instytut Podstaw Informatyki PAN, 2011.

- Kilgarriff, Adam, and Baisa, Vít, and Bušta, Jan et al. "The Sketch Engine: ten years on." *Lexicography. Journal of ASIALEX*, 2014, vol. 1, no. 1: 7–36, <<https://link.springer.com/article/10.1007/s40607-014-0009-9>> (11.10.21).
- Klimova, Marina V., and Turko, Ul'jana I., and Abreimova, Galina N. "O znachenii slova *kontent* v yazykovom soznanii molodezhi". *Nauchnyy dialog*, 2021, no. 7: 91–107, <https://cyberleninka.ru/article/n/o-znachenii-slova-kontent-v-yazykovom-soznanii-molodezhi> [Климова, Марина В., and Турко, Ульяна И., and Абреимова, Галина Н. "О значении слова *контент* в языковом сознании молодежи." *Научный диалог*, 2021, no. 7: 91–107, <<https://cyberleninka.ru/article/n/o-znachenii-slova-kontent-v-yazykovom-soznanii-molodezhi>> (21.10.21)].
- "Krasnyy Oktyabr'": *put' ot konditerskoy fabрики do art-klastera*, 30.09.2011, <<https://reality.ria.ru/20110930/396734786.html>> [*«Красный Октябрь»: путь от кондитерской фабрики до арт-кластера*, 30.09.2011, <<https://reality.ria.ru/20110930/396734786.html>> (11.10.21)].
- Kuznetsov, Sergey A. (Ed.). *Bol'shoy tolkovyy slovar' russkogo yazyka*. Sankt-Peterburg: Norint, 2014 (1998), <<http://www.gramota.ru/slovari/info/bts/>> [Ed. Кузнецов, Сергей А. Ed. *Большой толковый словарь русского языка*. СПб.: Норинт, 2014 (1998), <<http://www.gramota.ru/slovari/info/bts/>> (11.10.21)].
- Lewandowska-Tomaszczyk, Barbara, and Bańko, Mirosław, and Górski, Rafał L., and Łaziński, Marek, and Pezik, Piotr, and Przepiórkowski, Adam. "Narodowy Korpus Języka Polskiego: geneza i dzień dzisiejszy." *Narodowy Korpus Języka Polskiego*. Eds. Przepiórkowski, Adam, and Bańko, Mirosław, and Górski, Rafał L., and Lewandowska-Tomaszczyk, Barbara. Warszawa: Wydawnictwo Naukowe PWN, 2012: 3–10.
- Lewandowska-Tomaszczyk, Barbara. Ed. *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2005.
- Ministerstwo Spraw Zagranicznych Rosji, <<https://www.mid.ru/ru/home>> (11.10.21).
- Morkovkin, Valerij V., and Bogachëva, Galina F., and Lutskaya, Natal'ya M. *Bol'shoy universal'nyy slovar' russkogo yazyka*. Ed. Morkovkin, Valerij V. Moskva: "AST-PRESS KNIGA", "Slovari XXI veka," 2016 [Морковкин, Валерий В., and Богачёва, Галина Ф., and Луцкая, Наталья М. *Большой универсальный словарь русского языка*. Морковкин, Валерий В. (Ed.). Moskva: "АСТ-ПРЕСС КНИГА", "Словари XXI века", 2016].
- Narodowy Korpus Języka Rosyjskiego*, <<https://ruscorpora.ru/new/>> (11.10.21).
- Pawłowski, Adam. "Metodologiczne podstawy wykorzystania słowników frekwencyjnych w badaniu językowego obrazu świata." *Przestrzeń w językowym obrazie świata*. Eds. Pajdzińska, Anna, and Krzyżanowski, Piotr. Lublin, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 1999: 81–99.
- Plungyan, Vladimir A. "Korpus kak instrument i kak ideologiya: o nekotorykh urokakh sovremennoy korpusnoy lingvistiki." *Russkiy yazyk v nauchnom osveshchenii*, 2008, no. 2 (16): 7–20, <<http://rjano.ruslang.ru/ru/archive/2008-2/7-20>> [Плунгян, Владимир А. "Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики." *Русский язык в научном освещении*, 2008, no. 2 (16): 7–20, <<http://rjano.ruslang.ru/ru/archive/2008-2/7-20>> (11.10.21)].
- Plungyan, Vladimir A. "Zachem nuzhen Natsional'nyy korpus russkogo yazyka? Neformal'noye vvedeniye." *Natsional'nyy korpus russkogo yazyka: 2003–2005. Sbornik statey*. Ed. Plungyan, Vladimir A. Moskva: Indrik, 2005: 6–20, <<https://>

POLIFUNKCYJNOŚĆ KORPUSU JĘZYKOWEGO...

- ruscorpora.ru/new/sbornik2005/02plu.pdf> [Плунгян, Владимир А. “Зачем нужен Национальный корпус русского языка? Неформальное введение.” *Национальный корпус русского языка: 2003–2005. Сборник статей.* Плунгян, Владимир А. (Ed.). Москва: Индрик, 2005: 6–20, <<https://ruscorpora.ru/new/sbornik2005/02plu.pdf>> (11.10.21)].
- Severskaya, Ol'ga. *Govorim po-russki s Ol'goy Severskoy.* Moskva: Slovo/Slovo, 2004 [Северская, Ольга. *Говорим по-русски с Ольгой Северской.* Москва: Слово / Slovo, 2004].
- Sichinava, Dmitriy V. “Natsional'nyy korpus russkogo yazyka: ocherk predystorii.” *Natsional'nyy korpus russkogo yazyka: 2003–2005. Sbornik statey.* Ed. Plungyan, Vladimir A. Moskva: Indrik, 2005: 21–30, <<https://ruscorpora.ru/new/sbornik2005/03sitch.pdf>> [Сичинава, Дмитрий В. “Национальный корпус русского языка: очерк предыстории.” *Национальный корпус русского языка: 2003–2005. Сборник статей.* Ed. Плунгян, Владимир А. Москва: Индрик, 2005: 21–30, <<https://ruscorpora.ru/new/sbornik2005/03sitch.pdf>> (11.10.21)].
- Shvedova, Nataliya Yu. Ed. *Tolkovyy slovar' russkogo yazyka s vklucheniym svedeniy o proiskhozhdenii slov.* Moskva: Izdatel'skiy tsentr “Azbukovnik”, 2008 [Шведова, Наталия Ю. Ed. *Толковый словарь русского языка с включением сведений о происхождении слов.* Москва: Издательский центр “Азбуковник”, 2008].
- Sketch Engine access funded by ELEXIS: 2018–2022, <<https://www.sketchengine.eu/elexis/>> (11.10.21).
- Slovar' Akademii Rossiyskoy*, Chast' 1, ot A do G. Sankt-Peterburg: Imperatorskaya Akademiya nauk, 1789, <http://etymolog.ruslang.ru/doc/SAR1_A-G.pdf> [*Словарь Академии Российской: Словарь Академии Российской, Часть 1, отъ А до Г.* Санкт-Петербург: Императорская Академия наук, 1789, <http://etymolog.ruslang.ru/doc/SAR1_A-G.pdf> (11.10.21)].
- Tret'yakov, Vitaliy. “Diagnoz: upravlyayemaya demokratiya.” *Nezavisimaya gazeta*, 13.01.2000, <https://www.ng.ru/editor/2000-01-13/1_diagnos.html> [Третьяков, Виталий. “Диагноз: управляемая демократия.” *Независимая газета*, 13.01.2000, <https://www.ng.ru/editor/2000-01-13/1_diagnos.html> (11.10.21)].
- Ufimtseva, Natal'ya V. “Yazykovaya kartina mira kak predmet psikholingvisticheskogo issledovaniya.” *Język i metoda*, 2. Eds. Szumska, Dorota, and Ozga, Krzysztof. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2015 (published online: 2016): 249–256, <<https://www.ejournals.eu/Language-and-Method/2015/2015/art/6683>> [Уфимцева, Наталья В. “Языковая картина мира как предмет психолингвистического исследования.” *Język i metoda*, 2. Eds. Szumska, Dorota, and Ozga, Krzysztof. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2015 (published online: 2016): 249–256, <<https://www.ejournals.eu/Language-and-Method/2015/2015/art/6683>> (21.10.21)].
- Vaulina, Yekaterina Yu. *Informatika. Tolkovyy slovar'.* Moskva: Izdatel'stvo Eksmo, 2005 [Ваулина, Екатерина Ю. *Информатика. Толковый словарь.* Москва: Издательство Эксмо, 2005].
- Waliński, Jacek. “Typologia korpusów oraz warsztat informatyczny lingwistyki korpusowej.” *Podstawy językoznawstwa korpusowego.* Eds. Lewandowska-Tomaszczyk, Barbara. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2005.
- Żmigrodzki, Piotr Ed. *Wielki słownik języka polskiego PAN*, <<https://wsjp.pl/haslo/podglad/31267/demokracja-ludowa/3021049/ustroj>> (11.10.2021).



DANIEL BORYSOWSKI

 ORCID: <https://orcid.org/0000-0001-6594-9047>

Uniwersytet Opolski

WYKORZYSTANIE KORPUSÓW ROSYJSKOJĘZYCZNYCH NEWSÓW INTERNETOWYCH NA POTRZEBY SYSTEMÓW AUTOMATYCZNEGO ROZPOZNAWANIA MOWY W OBSZARZE MONITORINGU MEDIÓW

THE USE OF RUSSIAN-LANGUAGE INTERNET NEWS CORPORA FOR THE PURPOSES
OF AUTOMATIC SPEECH RECOGNITION SYSTEMS IN THE AREA OF THE MEDIA MONITORING

The author of the article used open Internet-news corpuses (NewsRu and Taiga) to create N-gram language models for the needs of automatic speech recognition systems. The models were comprehensively evaluated (perplexity, WER, proper name recognition, comparison with the base model and Google ASR). The author also rescored N-gram models, using recursive neural networks. The effectiveness of the models was assessed by recognizing speech from the news channel Россия 24 (37 files with a total length of 1.5 hours were tested). The selection of test data is related to the main goal of the article — speech recognition for the needs of the so-called media monitoring.

1. WPROWADZENIE

Istota percepcji mowy jako problem badawczy interesowała naukowców od dziesięcioleci. W wieku XIX ojciec amerykańskiej psychologii, William James, stwierdził: „Gdy słuchamy czyjejs mowy, wiele z tego, co — jak nam się wydaje — słyszymy, pochodzi z naszej pamięci”¹. Dziś, szczególnie w odniesieniu do automatycznego rozpoznawania mowy, to zdanie nabiera nowego sensu, choć gdy zostało opublikowane, badania w tym obszarze dalekie były od stworzenia faktycznego urządzenia ASR (ang. *automatic speech recognition*). Pierwsze warte odnotowania postępy w instrumentalizacji teorii dotyczącej

¹ W. James, *Talks to Teachers on Psychology: And to Students on Some of Life's Ideals*, Holt, New York, 1889, s. 159. Tłum. — D.B.

odbierania i przetwarzania mowy przez maszyny przypadają na połowę lat 50. XX wieku, kiedy w Bell Labs stworzono analogowy system do rozpoznawania cyfr wypowiedzianych przez człowieka do aparatu telefonicznego². W kolejnych dziesięcioleciach metodologia i technologia ASR rozwijane były dzięki postępom w różnych obszarach komputerowego przetwarzania danych³. Spośród ośmiu kluczowych obszarów (technologii), wymienionych przez Douglasa O’Shaughnessy’ego, z punktu widzenia niniejszego tekstu szczególne znaczenie ma *m o d e l o w a n i e j ę z y k a n a t u r a l n e g o* (na poziomie zarówno akustycznym, jak i graficznym⁴).

Obecnie trudno wskazać wyraźne granice między naukowym badaniem ASR, wykorzystaniem ASR w procesach biznesowych (telefoniczna obsługa klientów z wykorzystaniem tzw. botów głosowych, kompleksowa analiza językowa rozmów konsultantów *call centre* z klientami i in.), czy też wykorzystaniem ASR w życiu codziennym (wyszukiwanie głosowe, głosowe zarządzanie urządzeniami typu Smart, automatyczne generowanie napisów i in.)⁵. We wszystkich tych sferach działalności ludzkiej w pewnym stopniu korzysta się z najnowocześniejszych metodologii, zasobów oraz technologii; w ujęciu badawczym wykorzystanie takiego czy innego podejścia, zasobu czy technologii determinuje cel prowadzonych analiz. Jedną z możliwości zastosowania systemów ASR jest tzw. monitoring mediów, czyli rozpoznawanie mowy z programów radiowych i/ lub telewizyjnych, nadawanych na żywo i udostępnianych na przykład przez Internet. W niniejszym tekście prezentujemy proces przetwarzania rosyjskojęzycznych danych tekstowych (w różnym stopniu szczególności) właśnie na potrzeby modelowania języka oraz rozpoznawania mowy (na przykładzie programu informacyjnego *Россия 24*⁶).

2. MODELOWANIE JĘZYKA I AUTOMATYCZNE ROZPOZNAWANIE MOWY

Modelowanie języka w ostatnim dziesięcioleciu stało się kluczowym zagadnieniem w ramach przetwarzania języka naturalnego (ang. *na-*

² D. O’Shaughnessy, *Invited paper: Automatic speech recognition: History, methods and challenges*, „Pattern Recognition” 2008, nr 41, s. 2966–2967.

³ Por. tabelę 2. Tamże, s. 2967.

⁴ Mamy tutaj na myśli teksty pisane, prymarnie cyfrowe.

⁵ Niniejszy artykuł wpisuje się w nurt badań naukowych wspieranych przez sektor komercyjny, powstał bowiem dzięki współpracy autora z firmą VoiceLab.AI Sp. z o. o., która w ramach projektu nr POIR.01.01.01-00-1237/19 otrzymała częściowe dofinansowanie z funduszy Narodowego Centrum Badań i Rozwoju.

⁶ Nadawane na żywo za pośrednictwem strony <https://www.vesti.ru/> (10.10.2021).

tural language processing; dalej NLP) i jest wykorzystywane w wielu obszarach związanych, mówiąc najogólniej, z przetwarzaniem tekstów i wyzyskiwaniem z nich jakichś informacji (również jako jeden z subetapów automatycznego rozpoznawania mowy). Obecnie modele języka tworzy się w ramach tzw. uczenia maszynowego z wykorzystaniem sieci neuronowych, choć do różnych zadań z tego obszaru sięgano także po modele N-gramowe. Te ostatnie, zaliczane do modeli statystycznych i deterministycznych, znane były już od lat 50. XX wieku. W praktyce i na większą skalę stosowano je dopiero w kolejnych dekadach. Do dziś bywają wykorzystywane na potrzeby systemów ASR ze względu na łatwość ich budowania i implementowania⁷. Rozwój w sferze uczenia maszynowego doprowadził do wykorzystywania do zadań z obszaru NLP czy ASR modeli RNN, czyli modeli opartych na rekurencyjnych sieciach neuronowych (ang. *recurrent neural networks*). Stanowiły one albo podstawowe narzędzie, albo mechanizm dodatkowy, na przykład poprawiający rezultaty modeli N-gramowych. Ważnym wydarzeniem w sferze NLP było stworzenie algorytmu word2vec⁸, również wykorzystującego sieci neuronowe. W 2017 roku z kolei po raz pierwszy zaprezentowano nowy typ modelu — tzw. Transformer — obecnie najczęściej wykorzystywany typ modeli do zadań z obszaru przetwarzania języka naturalnego⁹.

W odniesieniu do ASR warto wymienić publikacje dotyczące historii tej technologii¹⁰. Tematykę ASR szczegółowo omawiają także podręczniki akademickie. Przykładowo w podręczniku Ivana Tam-

⁷ Więcej informacji na temat modeli N-gramowych por. choćby internetową (stale aktualizowaną i rozbudowywaną od roku 2008) publikację D. Jurafsky, J.H. Martin, *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Third Edition draft, 2021, s. 30–35 https://web.stanford.edu/~jurafsky/slp3/ed3book_sep212021.pdf (24.10.2021).

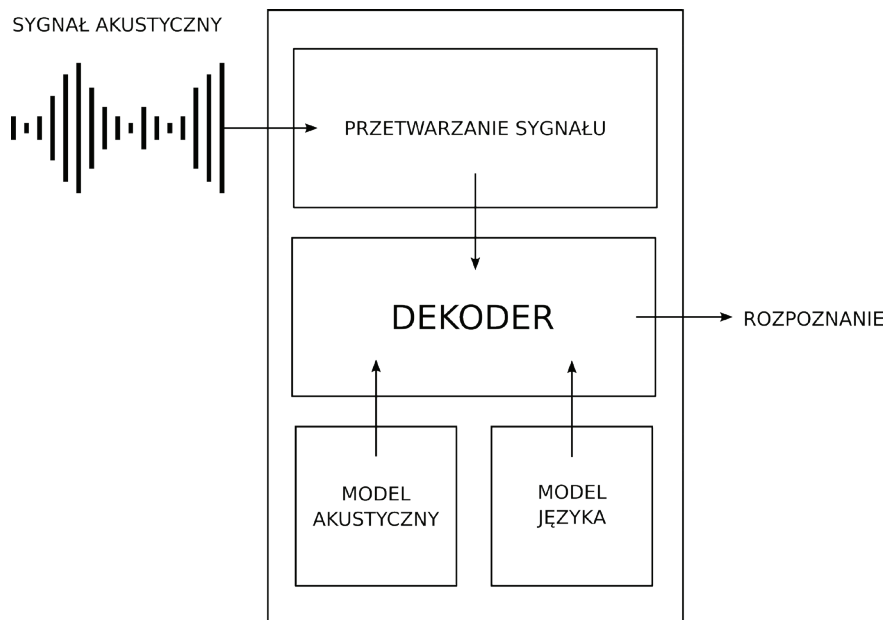
⁸ T. Mikolov i in., *Efficient Estimation of Word Representations in Vector Space*, 2013, <https://arxiv.org/abs/1301.3781v3> (20.10.2021) oraz T. Mikolov i in., *Distributed Representations of Words and Phrases and their Compositionality*, 2013, <https://arxiv.org/abs/1310.4546v1> (20.10.2021).

⁹ A. Vaswani i in., *Attention Is All You Need*, 2017, <https://arxiv.org/abs/1706.03762v5> (21.10.2021) oraz T. Wolf i in., *Transformers: State-of-the-Art Natural Language Processing*, 2020, <https://aclanthology.org/2020.emnlp-demos.6.pdf> (21.10.2021).

¹⁰ Zob. między innymi D. O’Shaughnessy, *Invited paper: Automatic speech recognition...*, B. Ziółko, M. Ziółko, *Przetwarzanie mowy*, Wydawnictwa AGH, Kraków, 2011 oraz И.Б. Тампель, *Автоматическое распознавание речи — основные этапы за 50 лет*, „Научно-технический вестник информационных технологий, механики и оптики” 2015, vol. 15, nr 6, s. 957–968.

pelya i Aleksey Karpova¹¹ kwestie automatycznego rozpoznawania mowy rozpatrywane są od etapu powstawania mowy w aparacie artykulacyjnym człowieka, aż po etap komputerowego przetwarzania sygnału dźwiękowego na tekst.

Zwiążłą informację dotyczącą modelowania języka oraz automatycznego rozpoznawania mowy wieńczymy schematem, który przedstawia działanie systemu ASR¹².



Rys 1. Podstawowy schemat systemu ASR

„Na wejściu” do systemu trafia sygnał akustyczny, na przykład w postaci pliku dźwiękowego. Odpowiednio przetworzony sygnał (do sekwencji obserwacji akustycznych) trafia do tzw. dekodera (programu dekodującego). Dzięki modelowi akustycznemu dekodery generuje sekwencję wyrazów odpowiadających otrzymanym reprezentacjom akustycznym. Model języka pomaga w wygenerowaniu naturalnego

¹¹ И.Б. Тампель, А.А. Карпов, *Автоматическое распознавание речи. Учебное пособие*, Университет ИТМО, Санкт-Петербург 2017.

¹² Schemat zaczerpnięty z R. Justo, O. Saz, A. Miguel, M.I. Torres, E. Lleida, *Improving Language Models in Speech-Based Human-Machine Interaction*, „International Journal of Advanced Robotic Systems” 2013, Vol. 10 (87), s. 1–11. Wykonanie — D.B.

tekstu w języku wyjściowym¹³. Jednym z kluczowych elementów skutecznego rozpoznawania mowy są tzw. słowniki¹⁴. Słownik to w istocie plik tekstowy, w którym każdy wyraz zajmuje jedną linię i jest podany wraz z jego postacią wyjściową oraz odpowiadającą mu fonemizacją. System ASR rozpoznaje tylko te wyrazy, które znajdują się w słowniku. Dla systemu korzystającego z modelu akustycznego literowego (więcej o tym zob. 4. Opis bazy materiałowej) przykładowe pozycje w słowniku przyjmą następującą postać:

```
автоматическое [автоматическое] а в т о м а т и ч е с к о е
распознавание [распознавание] р а с п о з н а в а н и е
речи [речи] р е ч и
антона [Антонa] а н т о н а
```

Wyrazy w słowniku — po lewej stronie — mają zazwyczaj taką samą postać, jak wyrazy w tekstowym korpusie, wykorzystanym do modelu (jeśli sprowadziliśmy wszystkie znaki alfabetu do małych liter). Słowa w nawiasach kwadratowych odpowiadają postaci wyjściowej z systemu (tak zostaną zapisane na przykład w pliku wyjściowym, wyświetlone na ekranie itp.). Pojedyncze litery po stronie prawej odpowiadają reprezentacjom akustycznym, które dekodery z rysunku 1 zamienia na wyrazy. System działa, jeśli model akustyczny został wytrenowany również przy użyciu słownika o takiej postaci (więcej o tym zob. 4. Opis bazy materiałowej).

3. MODELOWANIE JĘZYKA ROSYJSKIEGO I ASR DLA JĘZYKA ROSYJSKIEGO — ZARYS PROBLEMÓW

Język rosyjski z kilku powodów stanowi wyzwanie w odniesieniu do jego modelowania (warstwy tekstowej), a także rozpoznawania (warstwy akustycznej). W warstwie tekstowej wyzwania te wiążą się z przynależnością tego języka do grupy języków fleksyjnych¹⁵. Proble-

¹³ Tamże, s. 2.

¹⁴ Są one wkompileowane w model języka.

¹⁵ W przypadku licznych zadań z obszaru NLP przed procesem modelowania języka stosuje się tzw. lematyzację, czyli sprowadzenie wszystkich wyrazów do ich formy podstawowej. Teksty wykorzystane do tworzenia modeli akustycznych oraz modeli języka na potrzeby systemów ASR nie podlegają lematyzacji, aczkolwiek bywa ona stosowana do tzw. modeli dwuskładnikowych, w których równolegle z modelowaniem sekwencji wyrazów modelowane są także sekwencje właściwości gramatycznych tych wyrazów (por. A. Karpov, K. Markov, I. Kipyatkova, D. Va-

matyczna jest też duża dowolność szyku wyrazów w języku rosyjskim, co prowadzi do zamodelowania wielu „nadmiarowych” kontekstów, zmniejszając tym samym skuteczność modelu i zwiększając jego rozmiar. Por. zdania:

- [1] Вчера он был на работе.
- [2] Он вчера был на работе.
- [3] Он был вчера на работе.
- [4] Он на работе был вчера.
- [5] Он был на работе вчера.
- [6] На работе он был вчера.
- [7] Вчера на работе был он.
- [8] На работе вчера был он.

Dla językoznawcy różnica między tymi zdaniami jest oczywista – choćby ze względu na ich właściwości prozodyczne, wynikające z ich struktury tematyczno-rematycznej. Prawdopodobnie każda osoba władająca językiem rosyjskim bez problemu wychwyci wszelkie niuanse znaczeniowe zawarte w zdaniach [1]–[8]. Dla osób tych będzie oczywiste, że zdania [1]–[3] mówią o tym, że wczoraj pewien „on” był w pracy (a nie na zwolnieniu, urlopie itp.); zdania [4]–[6] informują z kolei o tym, że tenże „on” w pracy był wczoraj (a nie przedwczoraj), zdania [7]–[8] natomiast podkreślają, że w pracy był „on” (a nie kto inny)¹⁶. Dla modelu języka, którego źródłem jest tekst pisany (cyfrowy), różnicę stanowi jedynie szyk. Oczywiście nie każda z powyższych sekwencji wyrazów zostanie zamodelowana równie dobrze, co zależy od liczby ich wystąpień w przetwarzanym korpusie tekstowym¹⁷.

Inny problem w warstwie tekstowej (ortograficznej) stanowi dywiz. Z jednej strony istnieją w języku rosyjskim wyrazy o stałej lub względnie stałej pisowni z dywizem, na przykład zaimki nieokreślone что-то,

zhenina, A. Ronzhin, *Large vocabulary Russian speech recognition using syntactico-statistical language modeling*, „Speech Communication” 2013, Vol. 56, s. 213–228 oraz I. Kipyatkova, A. Karpov, *Study of Morphological Factors of Factored Language Models for Russian ASR*, w: A. Ronzhin i in. (eds.), *Speech And Computer*, Springer, Switzerland 2014, s. 451–458.

¹⁶ Subtelne różnice semantyczne między poszczególnymi zdaniami występują także w ramach trójek [1]–[3] i [4]–[6], a także w ramach dwójki [7]–[8], ale wymagałoby to precyzyjnego dowodzenia, które dla głównego wywodu nie jest niezbędne.

¹⁷ Odnosimy się tutaj głównie do modeli N-gramowych oraz RNN-owych. Nowsze typy modeli, np. modele Transformerowe (zob. przypis 9), wykorzystują mechanizm tzw. uwagi, i w pewnym stopniu potrafią wychwytywać zawiloci tematyczno-rematycznej struktury zdań.

кому-то, где-либо, кое-кто, przysłówki по-русски, по-моему, czy też połączenia typu чёрно-белый, культурно-просветительский, z drugiej zaś — szczególnie w potocznej odmianie ruszczyzny — takie dwukomponentowe twory, jak я-то, работаешь-то, успел-таки, принёс-таки i in. Łączliwość wyrazów z partykułą -to jest w zasadzie nieograniczona (czy też ograniczona jedynie przez uzus). Posiadanie wszystkich możliwych wariantów takich połączeń w słowniku znacznie zwiększy jego rozmiar (por. проблема-то, проблемы-то, проблем-то i in.) oraz może negatywnie wpłynąć na rozpoznanie w niektórych kontekstach.

Ostatnim problemem w warstwie tekstowej jest litera ë, choć stanowi ona także problem w warstwie akustycznej. Jednym z rozwiązań tego problemu jest zastępowanie litery ë literą e i złożenie kwestii dwoistej interpretacji akustycznej tego samego znaku graficznego na karb modelu akustycznego. Na poziomie analizy korpusów również wydaje się to uzasadnione, ponieważ sami Rosjanie raczej nie stosują litery ë w piśmie bądź stosują ją tylko w przypadkach kontekstowo niejednoznacznych (z reguły w tekstach literackich czy publicystycznych). Takie rozwiązanie zastosowaliśmy w naszych eksperymentach, co oznacza, że dla reprezentacji wyrazów всё i все w naszym słowniku znajduje się tylko jedna pozycja o postaci „все [все] в с е”, i to model akustyczny „wie”, że fonemowi e („fonemowi-literze” w literowym wariacie modelu) mogą odpowiadać w tym wypadku dwie wartości akustyczne.

Wydaje się, że większe wyzwanie stanowi warstwa akustyczna języka rosyjskiego. Ruchomy i swobodny akcent oraz związana z nim redukcja samogłosek w pozycji nieakcentowanej to główny problem systemów ASR służących rozpoznawaniu tego języka. Wyżej przedstawiliśmy fragment słownika, w którym zastosowano tzw. wariant literowy fonemizacji. W praktyce oznacza to znacznie łatwiejsze przygotowanie słownika na potrzeby ASR, lecz jednocześnie bardziej złożony proces tworzenia modelu akustycznego¹⁸. Wariant fonemowy tworzony jest na podstawie reguł, o ile takie w danym języku istnieją. W wypadku języka rosyjskiego — tak, wartości fonetyczne samogłosek są jednak uzależnione od pozycji akcentu w wyrazie. Rozwiązanie

¹⁸ Modele akustyczne trenuje się na parach plik dźwiękowy + plik tekstowy (z dokładnym zapisem tego, co zostało wypowiedziane w pliku dźwiękowym). W procesie treningu również stosuje się słowniki, które wiążą odpowiednie sekwencje fonemów i odpowiadające im wyrazy z warstwą akustyczną pochodzącą z pliku dźwiękowego.

stanowią programy zwane akcentorami statystycznymi¹⁹, które pozwalają na otrzymanie fonemizacji na przykład o takiej postaci²⁰:

автоматическое [автоматическое] A F T A M A To Io TSHo I S K A Jo I
 распознавание [распознавание] R A S P A Z N A V Ao No I Jo I
 речи [речи] Ro Eo TSHo I
 антона [Антонa] A N T Oo N A

Ostatnią kwestią wartą omówienia jest twarda bądź miękka wymowa wyrazów z literą „e”. Wyraz секс powinien być wymawiany twardo, podobnie wyrazy фонетика, энергия, компьютер i wiele innych. W niektórych przypadkach dopuszcza się oba warianty — miękki i twardy (por. na przykład wyraz бассейн), choć w odniesieniu do działania systemów ASR i tak istotne jest to, w jaki sposób ktoś coś wypowie, a nie to, jaką wymowę sugerują słowniki ortoepiczne (jaka jest przyjęta norma). Wspomniany w przypisie 19. program russian_g2p z tymi wypadkami sobie nie radzi i konsekwentnie oferuje fonemizację ze zmiękczeniem, co być może statystycznie jest nieistotne i nie wpłynie na błędne rozpoznanie tych wyrazów. W ramach wspólnych działań z firmą VoiceLab.AI (głównie z zespołem zajmującym się modelami akustycznymi) testowaliśmy różne rozwiązania, by ostatecznie pozostać przy wariacie literowym słownika (z fonemizacją zapisaną cyrylicą bez jakichkolwiek dodatkowych oznaczeń)²¹.

4. OPIS BAZY MATERIAŁOWEJ

Na początku tej części musimy wyjaśnić, co rozumiemy pod pojęciem „korpus”. Nie chodzi tu bowiem o typowy zbiór wyselekcjonowanych tekstów, zazwyczaj anotowanych, wyposażony w mniej lub bardziej zaawansowane narzędzie wyszukiwujące. W takich przypadkach dostęp do samych składowych korpusu jest ograniczony (teksty korpusu wyświetlane są fragmentarycznie, odpowiednio do określonego w kwe-

¹⁹ Zob. O. Yakovenko, I. Bondarenko, M. Borovikova, D. Vodolazsky, *Algorithms for automatic accentuation and transcription of Russian texts in speech recognition systems*, w: A. Karpov, O. Jokisch, R. Potapova (red.), *Speech And Computer...*, s. 768–777. Por. również https://github.com/nsu-ai/russian_g2p (15.10.2021).

²⁰ Znak zera przy samogłoskach oznacza akcent, natomiast przy spółgłoskach — obecność bezpośrednio po nich samogłoski zmiękczej.

²¹ Zastosowanie takiego rozwiązania wymaga jednak treningu modelu akustycznego na znacznie większej ilości danych — maksymalnie zróżnicowanych pod względem zarówno jakości, jak i przynależności gatunkowej.

rendzie korpusowej kontekstu). Z punktu widzenia przetwarzania języka naturalnego, a tym samym modelowania języka, potrzebne są po prostu dane tekstowe, na przykład w postaci pojedynczego pliku, którego kolejne linie stanowią swego rodzaju rekordy (każda linia to przykładowo jeden autonomiczny tekst, jedno zdanie, kilka zdań)²². To, jakiej postaci dane stosowane są do konkretnego zadania NLP, zależy od jego celu. Cel taki może stanowić klasyfikacja tekstów, generowanie słów kluczowych, tworzenie streszczeń, odtwarzanie interpunkcji, korekta rozpoznania OCR (ang. *optical character recognition*) i wiele innych²³. W modelowaniu języka pod kątem automatycznego rozpoznawania mowy przez korpus zwykle rozumie się zbiór (kolekcję) tekstów w formie plików tekstowych, takich, które można swobodnie przetworzyć w całości.

Do prezentowanego tutaj eksperymentu wykorzystaliśmy dwie ogólnodostępne kolekcje:

1) Korpus NewsRu (ponad 2,5 mln rosyjskojęzycznych newsów internetowych)²⁴,

2) Korpus Taiga, a dokładniej jego subkorpus, również dot. newsów internetowych (ponad 470 tys. tekstów)²⁵.

Pierwsza kolekcja gromadzi newsy internetowe z czterech popularnych serwisów informacyjnych (Вести, Интерфакс, Лента

²² Inną możliwością jest przechowywanie każdego tekstu/zdania w oddzielnych plikach, jeszcze inną — stosowanie różnych formatów plików służących do przechowywania danych, np. w układzie kolumnowym (TSV, CSV, CoNLL-X/CONLL-U) lub w postaci tzw. obiektów (np. JSON). W niniejszym tekście określeń „korpus tekstów”, „kolekcja tekstów”, „zbiór tekstów” będziemy używać zamiennie, jednak zawsze w odniesieniu do danych nieanotowanych, bez interfejsu graficznego, czy też bez „wbudowanego” graficznego narzędzia przeszukującego.

²³ Por. zadania w ramach kampanii ewaluacyjnej PolEval — <http://poleval.pl/> (15.20.2021). Opis jednego z nowszych i wysoce skutecznych modeli stosowanych obecnie w tego typu zadaniach (tzw. modelu T5), zob. C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P.J. Liu, *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*, „Journal of Machine Learning Research” 2020, vol. 21, s. 1–67.

²⁴ D. Borysowski, *Web crawling dla celów lingwistycznych. Wybrane aspekty gromadzenia i analizy danych tekstowych na przykładzie rosyjskojęzycznych newsów internetowych*, „Prace Językoznawcze” 2021, vol. XXIII/3, s. 87–104. Zob. także <https://osf.io/697md/> (10.10.2021).

²⁵ T. Shavrina, O. Shapovalova, *To the Methodology of Corpus Construction for Machine Learning: „Taiga” Syntax Tree Corpus and Parser*, w: V.P. Zakharov (red.), *Proceedings of the International Conference „Corpus Linguistics–2017”*, 2017, St. Petersburg, s. 78–84. Zob. także https://tatianashavrina.github.io/taiga_site/ (10.10.2021).

i Фонтанка) z lat 1999–2019. Druga — dane z serwisów Интерфакс, Лента і Фонтанка, a dodatkowo także z serwisu Комсомольская правда, w różnych proporcjach obejmując lata 2010–2017. Konstrukcja każdego korpusu jest nieco inna. Korpus NewsRu dostępny jest jako cztery odrębne pliki w formacie JSON. Fragment takiego pliku prezentujemy poniżej.

```
{
  "metadata": {
    "url": "https://lenta.ru/news/2019/12/24/oprovergli/",
    "date": „2019-12-24”,
    "time": "02:40:00",
    "category": "Бывший СССР",
    "subcategory": "Украина",
    "title": "СБУ прокомментировала сообщение о расстреле офицеров в Закарпатье",
    "tags": ["Прибалтика", "Белоруссия", "Молдавия", "Закавказье", "Казахстан", "Средняя Азия"]
  },
  "text": "Пресс-секретарь Службы безопасности Украины (СБУ) Елена Гитлянская опровергла информацию о расстреле оперативной группы в Закарпатье. Об этом она написала в Facebook. «Уважаемые журналисты, не распространяйте фейк о \”расстреле группы СБУ в Закарпатье\”. Это полный бред!» — говорится в сообщении представителя СБУ. О попытках и убийстве четырех сотрудников спецслужбы ранее, 23 декабря, сообщило украинское издание «Страна.ua». В публикации говорилось, что офицеров расстреляли в Закарпатье местные бандиты, которые занимались контрабандой. На момент публикации официального подтверждения информации не было.”
}
```

Przy użyciu odpowiedniego programu korpus ten można w dowolny sposób filtrować. Dzięki temu możliwe jest tworzenie rozmaitych podzbiorów danych (z tytułami, tagami, kategoriami, autorami), zliczanie liczby newsów opublikowanych w zadanym okresie (roku, miesiącu, dniu, przedziale godzinowym), zapisywanie wybranych elementów korpusu do plików wyjściowych (na przykład tylko tekstu zasadniczego newsów z wybranymi rekordami).

Korpus Taiga dostępny jest w dwóch postaciach. Jako zbiór plików tekstowych z tekstem zasadniczym newsów w formacie .txt — wraz z opisującymi go metaplikami (w formacie .csv), a także jako baza danych sqlite3. Pojedynczy rekord metapliku dotyczącego newsów z serwisu Лента prezentujemy poniżej w formie tabeli.

Tabela 1. Metaplik dotyczący pojedynczego newsa internetowego z zasobu Taiga (Лента)

Pole	Zawartość
segment	Lenta
Textid	20150806moonside
textname	НАСА показало обратную сторону Луны на фоне Земли
textrubric	Космос
Date	6 августа 2015
Time	07:09
Source	https://lenta.ru/news/2015/08/06/moonside/

Powższy metaplik powiązany jest z plikiem tekstowym 20150806moonside.txt. Por. jego fragment:

Космическому аппарату НАСА DSCOVR удалось получить редкий снимок обратной стороны Луны на фоне Земли — из точки Лагранжа системы Земля-Солнце (объекты, находящиеся в ней, одинаково притягиваются к обоим космическим телам). Об этом сообщается в пресс-релизе агентства. Исходную серию изображений камера Earth Polychromatic Imaging Camera (EPIC) на борту DSCOVR получила 16 июля 2015 года, с расстояния 1,5 миллиона километров от Земли [...].

Wybrane teksty z obu zasobów posłużyły nam do stworzenia dwóch modeli języka, które przetestowaliśmy z wykorzystaniem systemu ASR firmy VoiceLab.AI. Dane testowe stanowią fragmenty programów informacyjnych z kanału Россия 24 o łącznej długości ok. 1,5 godziny (37 plików). Dokonano ręcznej transkrypcji nagrań audio oraz — jako zadanie dodatkowe — ręcznej anotacji nazw własnych z podziałem na siedem kategorii: osoby (190 segmentów), wydarzenia (12 segmentów), produkty (28 segmentów), organizacje (43 segmenty), daty (21 segmentów), adresy (2 segmenty), lokalizacje (245 segmentów)²⁶. Podstawowym celem naszego badania było sprawdzenie, czy (i w jakim stopniu) wykorzystanie tych samych danych tekstowych z dwóch różnych źródeł wpływa na jakość modelu języka, a w efekcie — na wyniki automatycznego rozpoznawania mowy. Anotacja danych pozwoliła również na ustalenie skuteczności obu modeli w rozpoznawaniu nazw własnych.

²⁶ Modele, słowniki oraz dane testowe dostępne są w serwisie OSF na prawach licencji CC BY-NC 4.0 (<https://osf.io/a6hm5/>). Wśród danych tych czytelnik nie odnajdzie jedynie pliku z modelem akustycznym oraz plików wejściowych dekodera, stanowiących własność firmy VoiceLab.AI.

5. PROCES TWORZENIA MODELI JĘZYKA ORAZ ICH EWALUACJI

Opisane korpusy — dzięki metadansom — mogą być wykorzystywane do różnych zadań z obszaru NLP; nas interesowała jedynie treść zawartych w nich tekstów. Zbiór NewsRu jest kilkukrotnie większy od zbioru Taiga (jego subkorpusu newsów), dlatego do naszych eksperymentów wykorzystaliśmy wszystkie dokumenty tekstowe zbioru Taiga News (z wyłączeniem danych serwisu Комсомольская правда) oraz wszystkie odpowiadające im dokumenty ze zbioru NewsRu (rzecz jasna, z wyłączeniem danych serwisu Вести). Tę odpowiedniość ustaliliśmy dzięki identyczności adresów URL w metadansom obu zbiorów. W sumie dało to około 400 tysięcy pokrywających się „rekordów”. Tak przygotowane zasoby zostały podzielone na część treningową, walidacyjną oraz testową w proporcji 0,8/0,1/0,1. Dane treningowe wykorzystaliśmy do stworzenia modeli N-gramowych (a dokładniej — 3-gramowych) oraz modeli RNN²⁷. Te drugie miały posłużyć do rescoringu, czyli poprawy wyników uzyskanych z użyciem modeli bazowych²⁸. Proces podziału na tzw. zbiory TRAIN, VALID i TEST poprzedziła normalizacja korpusów, która oznaczała odpowiednio:

- 1) usunięcie znaków interpunkcyjnych (w tym zamianę dywizu na spację),
- 2) zamianę litery ë na literę e,
- 3) zamianę wielkich liter na małe,
- 4) zamianę cyfr i liczb na pomocniczy tag <NUMB>, który z założenia powinien pomóc w lepszym zamodelowaniu kontekstów z liczebnikami, datami i in.,
- 5) usunięcie znaczników HTML, adresów internetowych oraz adresów poczty elektronicznej,
- 6) usunięcie nie-cyrylicy.

Po zastosowaniu normalizacji przytaczany już fragment tekstu o działaniach organizacji NASA przyjął następującą postać:

космическому аппарату наса удалось получить редкий снимок обратной стороны луны на фоне земли из точки лагранжа системы земля солнце объ-

²⁷ Modele N-gramowe zbudowaliśmy narzędziem IRSTLM (zob. <https://hlt-mt.fbk.eu/technologies/irstlm>; 26.10.2021), modele RNN natomiast — narzędziem opartym na kodzie TensorFlow (zob. <https://www.tensorflow.org/>; 26.10.2021).

²⁸ Por. wyniki badań prowadzonych w tym kierunku: I. Kipyatkova, A. Karpov, *Recurrent Neural Network-based Language Modeling for an Automatic Russian Speech Recognition System*, w: S. Balandin i in. (red.), *Proceedings of AINL-ISMW FRUCT Conference*, St. Petersburgs 2015, s. 33–38.

екты находящиеся в ней одинаково притягиваются к обоим космическим телам об этом сообщается в пресс релизе агентства исходную серию изображений камера на борту получила <NUMB> июля <NUMB> <NUMB> года с расстояния <NUMB> <NUMB> миллиона километров от земли

Modele N-gramowe będziemy dalej nazywać — zgodnie z formatem, w jakim są zapisane — plikami ARPA²⁹, dodając prefiks danego zasobu, na przykład NewsRu ARPA oraz Taiga ARPA. Modele RNN będziemy odpowiednio identyfikować nazwami NewsRu RNN, Taiga RNN. Określenie „model języka” zwyczajowo stosuje się zarówno w odniesieniu do formatu ARPA, jak i w odniesieniu do modeli RNN, a także finalnego pliku wejściowego do dekodera, dlatego — w celu uniknięcia terminologicznych nieporozumień — wprowadzamy powyższe rozróżnienie. Dodatkowo gotowy system automatycznego rozpoznawania mowy będziemy oznaczać skrótem „ASR”. Podstawą w przeprowadzonym eksperymencie będą dla nas zatem wszystkie wymienione modele, a także systemy NewsRu ASR, Taiga ASR, punktem odniesienia natomiast — Base ASR³⁰ oraz Google ASR. Ewaluacja plików ARPA, modeli RNN oraz kompletnych systemów ASR odbywała się na dwóch poziomach:

- 1) Perplexity (ARPA i RNN) na danych VALID i TEST oraz na plikach tekstowych właściwego zbioru testowego (dalej TEST SET)³¹,
- 2) WER (ASR) na danych TEST SET³².

Wyniki analizy perplexity prezentujemy w tabeli 2. Tabela zawiera również informację o tzw. OOV (ang. *out-of-vocabulary*), czyli liczbie wyrazów z danych testowych i walidacyjnych, których nie ma w modelu (wyrażonej procentowo). Pogrubieniem oznaczamy naj-

²⁹ M. Federico, N. Bertoldi, M. Cettolo, *IRSTLM: an Open Source Toolkit for Handling Large Scale Language Models*, w: *Proceedings of Interspeech*, Brisbane 2008, s. 1618–1621.

³⁰ Base ARPA oraz Base RNN zostały stworzone z danych korpusu OpenSubtitles 2018 (<https://opus.nlpl.eu/>; 29.10.2021), wybranych danych z korpusu Taiga oraz danych komercyjnych kupionych przez firmę VoiceLab.AI.

³¹ Zob. D. Jurafsky, J.H. Martin, *Speech and Language Processing...*, s. 36–37. Mówiąc najogólniej, perplexity jest miarą określającą skuteczność przewidywania przez model języka danych testowych, których wcześniej „nie widział”, czyli nie został stworzony na ich podstawie. Im niższa wartość perplexity, tym lepszy model.

³² Od ang. *word error rate*. Jest to miara określająca, jaki procent rozpoznania stanowią błędy. Błędy te z kolei możemy podzielić na trzy kategorie: usunięcia (ang. *deletions*), wstawienia (ang. *insertions*) oraz podstawienia (ang. *substitutions*). Suma wszystkich usunięć, wstawień i podstawień obliczona poprzez porównanie pliku z dokładną transkrypcją oraz pliku z wyjściowym rozpoznaniem z dekodera stanowi wskaźnik WER.

lepszy wynik, przy czym bierzemy pod uwagę tylko modele NewsRu i Taiga.

Tab. 2. Perplexity dla modeli NewsRu, Taiga oraz Base (ARPA i RNN)

Dane	Modele	NewsRu		Taiga		Base	
		Typ modelu		Typ modelu		Typ modelu	
		ARPA	RNN	ARPA	RNN	ARPA	RNN
NewsRu	VALID	PP=411.71 OOV=2.66%	PP=204.76	PP=468.97 OOV=2.72%	PP=237.60	PP=420.42 OOV=2.40%	PP=330.64
	TEST	PP=522.36 OOV=3.03%	PP=234.75	PP=551.76 OOV=3.05%	PP=255.16	PP=425.58 OOV=2.30%	PP=356.69
Taiga	VALID	PP=159.81 OOV=2.39%	PP=199.95	PP=471.89 OOV=2.62%	PP=180.25	PP=505.71 OOV=2.66%	PP=302.59
	TEST	PP=579.77 OOV=3.55%	PP=238.67	PP=599.01 OOV=3.57%	PP=248.13	PP=520.51 OOV=2.84%	PP=354.20
TEST SET (Россия 24)		PP=2416.37 OOV=4.17%	PP=692.62	PP=2457.19 OOV=4.21%	PP=719.52	PP=1507.52 OOV=1.65%	PP=796.61

Bardzo niska wartość perplexity dotycząca modelu NewsRu ARPA oraz danych Taiga VALID najprawdopodobniej spowodowana jest wysokim stopniem pokrywania się tych danych z danymi TRAIN modelu NewsRu (podczas procesu przygotowywania korpusów do treningu zasoby dzielone są w sposób losowy). Dla pozostałych zasobów perplexity utrzymuje się na względnie stałym poziomie, co oznacza, że tę jedną wartość należałoby potraktować jako nieistotną.

Kolejnym etapem tworzenia systemu ASR jest opracowanie słownika. Listy słów do słowników są generowane automatycznie podczas procesu obróbki danych korpusowych. Przetwarza się je do list rangowych, aby możliwe było wyselekcjonowanie najczęstszych wyrazów oraz odrzucenie tych najrzadszych bądź po prostu niechcianych. Przygotowując dane do treningu modelu RNN (te same dane zostały potem wykorzystane do stworzenia plików ARPA), wyznaczyliśmy granicę powtórzeń wyrazów na dość wysokim poziomie 48, co dało w wypadku obu zasobów (NewsRu i Taiga) około 130 tysięcy wyrazów. Listy te zostały wykorzystane jako podstawa słowników. Dodatkowo z danych treningowych stworzyliśmy słowniki zawierające po ok. 300 tysięcy wyrazów oraz wszystkie wyrazy z danego zasobu (NewsRu — 642 tysiące, Taiga — 677 tysięcy), aby sprawdzić wpływ

rozmiaru słownika na WER. To oznacza, że do testów zbudowaliśmy sześć modeli (po trzy na podstawie każdego z zasobów z trzema różnymi słownikami).

Jednym z najważniejszych elementów systemu ASR jest model akustyczny. Do testów wykorzystaliśmy model literowy, wytrenowany na 2539 godzinach nagrań³³.

Etap testowania modeli na drugim poziomie — dotyczącym WER — odbywał się w kilku subetapach oraz dotyczył wyłącznie danych TEST SET, czyli 1,5 h nagrań z kanału informacyjnego Россия 24. Kolejność działań była następująca:

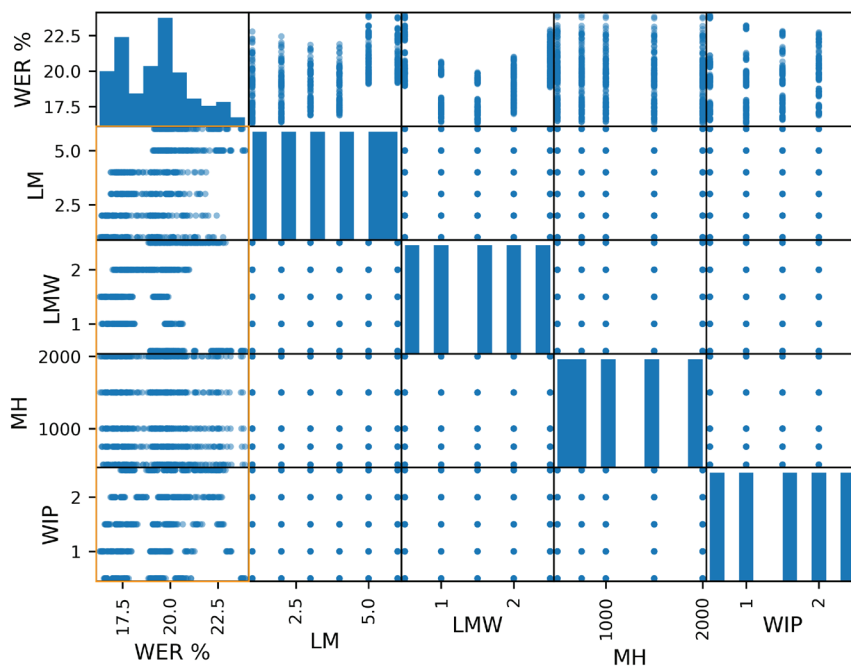
- 1) poszukiwanie najlepszej konfiguracji modelu języka / parametrów modelu akustycznego i dekodera / rozmiaru słownika (modele NewsRu ASR i Taiga ASR),
- 2) wybór najlepszej konfiguracji,
- 3) testy modelu Base ASR z użyciem kilku najlepszych wariantów konfiguracji,
- 4) testy modelu Google ASR (bez możliwości ingerencji w konfigurację, ale z różnymi wariantami normalizacji rozpoznań, zob. przypis 33),
- 5) rescoring modeli N-gramowych (ARPA) z użyciem modeli RNN,
- 6) testy modeli z rescoringiem dla wybranych konfiguracji.

Dodatkowym subetapem był wspomniany już test poprawności rozpoznawania nazw własnych. Wszystkie powyższe etapy ewaluacyjne zostały przeprowadzone w wariancie lowercase (czyli bez uwzględniania wielkości liter), choć nasze modele są przygotowane do wyświetlania rozpoznań z zachowaniem wielkich liter w jednowyrazowych nazwach własnych.

W pierwszym subetapie sprawdziliśmy 750 wariantów konfiguracji (po 125 dla każdego z sześciu modeli). Konfiguracje obejmowały

³³ Konkretniej są to dane Golos — 1221 godzin (<https://github.com/sberdevices/golos>; 28.10.2021), 5-procentowa próbka danych Open-STT — 657 godzin (https://github.com/snakers4/open_stt; 28.10.2021), kanał informacyjny Первый — 11 godzin (własne transkrypcje), kanał informacyjny Россия 24 — 9 godzin (dane nieużyte w danych TEST SET; własne transkrypcje), Common Voice — 104 godziny (<https://commonvoice.mozilla.org/pl>; 28.10.2021), Kaggle — 134 godziny (<https://www.kaggle.com/tapakah68/audio-dataset>; 28.10.2021), LibriSpeech — 93 godziny (<https://www.openslr.org/96/>; 28.10.2021), Radio Svoboda — 24 godziny (własne transkrypcje), Multilingual TEDx — 40 godzin (<https://www.openslr.org/100>; 28.10.2021), Caito — 46 godzin (<https://www.caito.de/>; 28.10.2021), VoxForge — 18 godzin (<http://www.voxforge.org/ru>; 28.10.2021) oraz inne dane firmy VoiceLab.AI (pozostałe 182 godziny).

takie parametry, jak LMW (ang. *language model weight*; waga modelu języka względem modelu akustycznego), MH (ang. *max hypotheses*; liczba hipotez z modelu akustycznego), WIP (ang. *word insertion penalty*; kara za wstawianie słów, ograniczająca rozbijanie wyrazów na mniejsze komponenty, jeśli te stanowią samodzielne wyrazy, na przykład *разводить/раз водить*). Poniżej prezentujemy wyniki ewaluacji³⁴.



Rys. 2. Zależność WER od rozmiaru słownika oraz parametrów LMW, MH, WIP

Rysunek 2 jasno pokazuje, że najmniejszy wpływ na jakość rozpoznań miał parametr MH, choć nieznacznie lepsze wyniki dawały jego wartości na poziomie 1500–2000. Najlepiej radziły sobie modele 1 i 2 (Taiga ASR ze słownikiem 677 tys. wyrazów oraz NewsRu ASR ze słownikiem 642 tys. wyrazów), choć trudno tu mówić o wyraźnej przewadze (w pierwszych 75 konfiguracjach – WER poniżej progu 17% – pojawiają się także modele 3 i 4 ze średnimi słownikami, a różnica między najlepszym a najgorszym wynikiem w tym przedziale to

³⁴ Są one także dostępne w formacie .xlsx na stronie <https://osf.io/a6hm5/> (29.10.2021).

0,68363 punktu procentowego). Najlepszy wynik WER modelu 5 to 19,08965% (315 pozycja spośród wszystkich konfiguracji; model NewsRu ASR), natomiast modelu 6 – 19,16753% (325 pozycja; model Taiga ASR). Zależność między rozmiarem słownika a wynikiem WER jest bardzo wyraźna. Z rysunku 2 wynikają także najlepsze warianty parametrów LMW i WIP (odpowiednio 1–1,5 i 1). Na 14 najlepszych wyników parametr WIP=1 miały wszystkie, parametr LMW=1,5 – 9 na 14. 5 pozostałych wyników dotyczyło LMW=1. Parametr MH, jak wspomnieliśmy, miał tutaj najmniejsze znacznie, gdyż z wyjątkiem jego wartości równej 500 pozostałe wystąpiły po trzy razy na 14. Najniższy WER odnotował model Taiga ASR – 16.31187% (także dwa kolejne wyniki: 16.32918% i 16.39841%). Model NewsRu ASR zajął czwarte miejsce z wynikiem 16.40706%³⁵. Warto zaznaczyć, że różnice na tym poziomie są raczej niezauważalne dla ludzkiego oka i nie sposób ich dostrzec podczas wzrokowej analizy tekstów rozpoznania. W tabeli 3 zamieszczamy listę dziesięciu najczęstszych wyrazów z kategorii usunięć, wstawień i podstawień dla obu porównywanych tutaj zasobów (NewsRu i Taiga) oraz modelu referencyjnego Base ASR. Wzięliśmy pod uwagę tylko najlepsze wyniki.

Tab. 3. Lista najczęstszych wyrazów z kategorii usunięć, wstawień i podstawień

Usunięcia			Wstawienia			Podstawienia		
NewsRu	Taiga	Base	NewsRu	Taiga	Base	NewsRu	Taiga	Base
1.973%	1.91243%	2.11146%	3.4441%	3.51333%	3.01142%	10.98996%	10.88612%	10.15922%
и (42)	и (43)	и (41)	в (12)	в (14)	в (13)	ну / но (9)	ну / но (8)	ни / не (5)
в (33)	в (34)	в (38)	и (11)	и (11)	и (12)	ни / не (6)	ни / не (7)	ну / но (5)
а (24)	а (25)	а (16)	на (6)	на (6)	на (7)	из / и (5)	нету / нет (4)	и / таки (4)
не (12)	с (11)	не (12)	это (6)	это (5)	а (5)	нету / нет (4)	в / и (3)	владимир / владимира (3)
с (12)	вот (10)	с (11)	но (4)	а (4)	но (5)	владимир / владимира (3)	владимир / владимир (3)	гуайдо / до (3)
вот (10)	не (9)	вот (7)	а (3)	за (4)	это (5)	которые / который (3)	из / и (3)	с / из (3)

³⁵ Wyniki te porównaliśmy z wynikami modeli referencyjnych: modelu Base ASR oraz Google ASR. Pierwszy – z parametrami MH=1500, LMW=1,5, WIP=1 – uzyskał wynik WER 15.2821%, drugi – 27.890% (teksty oryginalnych rozpoznania), 24.957% (teksty po zamianie dywizu na spację oraz litery ё na literę е), 22.975% (teksty po normalizacji liczb zapisanych cyframi do zapisu słownego; ASR Google generuje rozpoznania różniące się od przyjętej przez nas normy, dlatego wymagane było ich ujednolicenie przed obliczeniem WER).

WYKORZYSTANIE KORPUSÓW ROSYJSKOJĘZYCZNYCH...

у (9)	о (9)	на (6)	за (3)	о (3)	все (3)	не / они (3)	не / они (3)	что / чтобы (3)
на (8)	у (9)	о (6)	о (3)	с (3)	за (3)	с / из (3)	с / из (3)	шарыга / га (3)
о (8)	на (7)	у (6)	с (3)	больше (2)	больше (2)	эта / это (3)	эта / это (3)	эта / это (3)
к (7)	но (7)	как (5)	больше (2)	все (2)	был (2)	акции / акций (2)	академия / академии (2)	академия / академии (2)

W kategoriach usunięć oraz wstawień najczęściej pojawiają się krótkie wyrazy. Operowanie parametrem WIP prowadzi często do przesunięć pomiędzy tymi kategoriami błędów (im wyższy WIP, tym więcej usunięć, im niższy — tym więcej wstawień), jednak nie eliminuje problemu. Nawet odnalezienie optymalnej konfiguracji stanowi raczej kompromis między liczbą usunięć a wstawień. Jeśli chodzi o podstawienia, w tabeli 3 odnajdujemy co najmniej trzy ich przyczyny:

1) podobieństwo na poziomie fonetycznym związane głównie z redukcją samogłosek (por. не / ни, ни / не, нету / нет, это / эта, ну / но);

2) problemy z nazwami własnymi — na przykład z ich brakiem w słowniku, przez co model podstawia w ich miejsce wyrazy o zbliżonej reprezentacji akustycznej (por. гуайдо / до, шарыга / га poprzedzone innymi wyrazami zbliżonymi w warstwie dźwiękowej do гуай / шары);

3) ewentualne błędy w transkrypcjach, stanowiących podstawę porównania z rozpoznaniem (por. владимир / владимира, акции / акций, академия / академии).

Właściwie wszystkie opisane tutaj kategorie mogą dotyczyć po prostu nieprawidłowości w ręcznych transkrypcjach, które zawsze obarczone są błędem ludzkim. Problem z punktu 2) jest jednym z podstawowych wyzwań współczesnych systemów automatycznego rozpoznania mowy. Sprawdzenie jakości naszych modeli w tym obszarze było zresztą naszym celem dodatkowym. Nazwy własne z siedmiu zdefiniowanych przez nas kategorii obejmowały łącznie 6,72% tekstów transkrypcji. Oba modele poradziły sobie z tym zadaniem na zbliżonym poziomie skuteczności — model NewsRu w 74,68%, model Taiga — w 75,1% (przypomnijmy, że analiza dotyczyła wariantu lowercase). Wyżej wzmiankowaliśmy, że analizowane modele są przystosowane do wyświetlania jednowyrazowych nazw własnych. Aby sprawdzić ich działanie, wykonaliśmy jeszcze jeden dodatkowy test, tym razem w wariantcie uppercase. Poniżej prezentujemy tekst roz-

poznany przez oba porównywane tu systemy z użyciem największych słowników. Błędnie rozpoznane wyrazy oznaczamy odpowiednimi numerami w nawiasach kwadratowych — [1] NewsRu ASR: ogółem uppercase WER 17.70509% (w poniższym przykładzie — 11.88118%), [2] Taiga ASR: ogółem uppercase WER 17.67047% (w poniższym przykładzie — 11.88118%). Dla czytelności znakiem „/” oznaczamy granice zdań.

сенат США не смог одобрить законопроект по санкциям против Сирии и ее союзников Ирана и России / документ даже не дошел до голосования не набрав двух т р е т ь е й [1 i 2 zamiast trzeciej] голосов / рассмотрение перенесли / законопроект предусматривает санкции за военную или иную помощь правительству Сирии и даже за сотрудничество в производстве углеводородов / прописанные знакомые обвинения в адрес России к о т о р о е [1 zamiast которая] якобы мешала доставки гумпомощи что н а ш и [1 i 2 (наша) zamiast наше] Минобороны не раз о п р о в е р г а л а [1 i 2 zamiast опровергаю] / впрочем дело не в содержании бумаги а в американском ш а н д а у [1 i 2 zamiast шатдауне] н е к о т о р ы е [1 i 2 zamiast который] тянется с конца декабря / демократы в сенате не желают рассматривать никакие инициативы кроме тех что п о л о ж и т [1 i 2 zamiast положат] конец ш а т у н у [1 i 2 (шандау ну) zamiast шатдауну] и п о з в о л и т [1 i 2 zamiast позволят] правительству США получить финансирование и возобновить

Przedstawienie kompletnych wyników w tym miejscu nie jest możliwe, jednak już jeden przytoczony tekst obrazuje zarówno mocne, jak i słabe strony stworzonych przez nas modeli (a także ogólne problemy związane z rozpoznawaniem mowy w języku rosyjskim). Podstawowym problemem, wymienianym już wielokrotnie, jest redukcja rosyjskich samogłosek. Wartości fonetyczne zakończeń wyrazów — na przykład нарушение/нарушения, которое/которая i wielu innych — są zbliżone, co, na przykład w połączeniu z szybkim tempem mowy, znacznie utrudnia prawidłowe rozpoznanie. Być może słowniki stworzone na podstawie reguł g2p nie będą obciążone tą bliskością, jednak nie dysponujemy wariantem fonemowego modelu akustycznego, aby móc sprawdzić te różnice. Dysponujemy natomiast wynikami poprzedniego modelu akustycznego firmy VoiceLab.AI (158 godzin), który na danych Россия 24 osiągał WER na poziomie ok. 35%. Pokazuje to, jak kosztowne jest poprawianie wyników ASR (10 razy więcej godzin dało poprawę o ok. 18 punktów procentowych).

Innym problemem jest rozpoznawanie nazw własnych (ok. 75% skuteczności w wariancie lowrcase), choć w omawianym fragmencie tekstu wszystkie nazwy zostały rozpoznane prawidłowo. Nie został rozpoznany wyraz stosowany dla określenia wstrzymania prac ame-

rykańskiego rządu, czyli шатдаун (ang. *shutdown*). Jak widać, modele „próbują” jakoś tę kwestię rozwiązać, wstawiając najbardziej zbliżony wariant brzmieniowy (na podstawie wyrazów zawartych w słowniku). W słownikach znalazły się takie wyrazy jak шандай (od Бад-Шандау, miasto-uzdrowisko w Niemczech) oraz шатуна (dopełniacz wyrazu шатун), które „upodobniły” się do wyrazów шатдауне ([шандау + не]которые) i шатдауну (шатуны i [шандау + ну]). Pierwszy przypadek doprowadził do dwóch błędów – nierozpoznania wyrazu шатдаун oraz niepoprawnego rozpoznania wyrazu который.

Dwa ostatnie subetapy naszych eksperymentów to rescoring modeli N-gramowych oraz testy modeli po rescoringu. Po przeprowadzeniu kilku treningów oraz doborze parametrów uzyskaliśmy nieznaczną poprawę, dodatkowo tylko przy użyciu do rescoringu modelu Base RNN (co obrazuje tabela 4). NewsRu RNN oraz Taiga RNN dawały minimalną poprawę lub wręcz nieco pogarszały poziom WER³⁶.

Tab. 4. Wyniki rescoringu modeli N-gramowych

Model	WER	Usunięcia	Wstawienia	Podstawienia	Waga rescoringu
[1]	16.06957%	1.81724%	3.69505%	10.55729%	0.1
[2]	16.21668%	1.79128%	3.78159%	10.64382%	
[3]	16.36379%	1.84320%	3.72966%	10.79093%	
[4]	16.44168%	1.79993%	3.76428%	10.87747%	
[B]	14.9187%	1.84320%	3.22776%	9.84770%	

Model [1] to NewsRu ARPA po rescoringu modelem Base RNN, model [2] – to Taiga ARPA po rescoringu tym samym modelem. Modele [3] i [4] to odpowiednio modele NewsRu ARPA i Taiga ARPA po rescoringu odpowiadającymi im modelami RNN. Powyższe wyniki mogą świadczyć o tym, że dla większej poprawy należałoby wykorzystać więcej danych, ale być może nie danych o charakterze ogólnym (por. podstawę modelu Base RNN w przypisie 30), lecz wyłącznie danych domenowych (na przykład publicystycznych). Sam model Base RNN nie wyczerpuje możliwości treningowych, o czym świadczy wysokie perplexity uzyskane na danych testowych tego modelu (ponad

³⁶ Wyjściowy WER modelu NewsRu ARPA wyniósł 16.40706%, natomiast Taiga ARPA – 16.31187%. Dane treningowe oraz pliki konfiguracyjne udostępniamy w serwisie OSF (<https://osf.io/a6hm5/>; 30.10.2021).

780), chociaż dla wyniku WER 15,2821%, uzyskanego na modelu Base ARPA, po rescoringu modelem RNN wytrenowanym na tych samych danych uzyskaliśmy spadek WER do 14,91866% (zob. model [B] w tabeli 4). Rescoring modeli N-gramowych NewsRu ARPA i Taiga ARPA z wykorzystaniem tych samych danych do modelu RNN nie przyniósł zadowalającej poprawy³⁷.

6. PODSUMOWANIE

Skuteczność systemów ASR jest coraz większa, a na ogólny jej wzrost składa się wiele czynników. Na poziomie akustycznym bada się możliwości ograniczenia ilości danych dźwiękowych, których dokładne transkrybowanie jest bardzo kosztowne³⁸. Zarówno na poziomie akustycznym, jak i na poziomie graficznym (tekstowym) stosuje się tzw. augmentację danych, czyli wykorzystywanie nieco zmienionych kopii posiadanych danych (lub danych generowanych automatycznie na podstawie posiadanych)³⁹. Zapewne kolejnym krokiem w uzyskaniu poprawy w naszym przypadku będzie wykorzystanie któregoś z powyższych podejść. Pierwszym zadaniem w ramach monitoringu mediów może być lepszy dobór danych na poziomie korpusowym. Oba przedstawione tutaj zbiory tekstów (NewsRu i Taiga) dają takie możliwości dzięki bogatemu metaopisowi (tytuły, kategorie, tagi itp.). Problemem może być rozbieżność funkcjonalna między tekstem pisanym a mówionym, dlatego wykorzystanie również techniki wav2vec mogłoby przynieść poprawę rezultatów, stanowi to jednak kolejny problem badawczy, wymagający omówienia w innym miejscu.

³⁷ W jednym ze znanych nam eksperymentów dotyczących tej samej kwestii badacze uzyskali poprawę o 3,67 punktu procentowego, co oznaczało obniżenie WER z poziomu 26,54% do poziomu 22,87%. Por. I. Kipyatkova, A. Karpov, *Recurrent Neural Network-based Language Modeling...*, s. 37.

³⁸ Por. zastosowanie algorytmu wav2vec, który pozwala na wykorzystanie od zaledwie kilku do kilkunastu godzin nagrań z transkrypcjami i trenowanie modelu akustycznego wyłącznie na danych dźwiękowych, S. Schneider, A. Baevski, R. Collobert, M. Auli, *wav2vec: Unsupervised Pre-training for Speech Recognition*, 2019 (<https://arxiv.org/abs/1904.05862v4>; 31.10.2021).

³⁹ A. Laptev, R. Korostik, A. Svischev, A. Andrusenko, I. Medennikov, S. Rybin, *You Do Not Need More Data: Improving End-To-End Speech Recognition by Text-To-Speech Data Augmentation*, 2020 (<https://arxiv.org/abs/2005.07157v2>; 31.10.2021).

REFERENCES

- Borysowski, Daniel. "Web crawling dla celów lingwistycznych. Wybrane aspekty gromadzenia i analizy danych tekstowych na przykładzie rosyjskojęzycznych newsów internetowych." *Prace Językoznawcze*, 2021, Vol. XXIII/3: 87–104.
- Federico, Marcello, and Bertoldi, Nicola, and Cettolo, Mauro. *IRSTLM: an Open Source Toolkit for Handling Large Scale Language Models. Proceedings of Interspeech*, Brisbane, 2008: 1618–1621.
- James, William. *Talks to Teachers on Psychology: And to Students on Some of Life's Ideals*. New York: Holt, 1889.
- Jurafsky, Dan, and Martin, James H. *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Third Edition draft, 2021 <https://web.stanford.edu/~jurafsky/slp3/ed3book_sep212021.pdf>.
- Justo, Raquel, and Saz, Oscar, and Miguel, Antonio, and Torres, M.I., and Lleida, Eduardo. "Improving Language Models in Speech-Based Human-Machine Interaction." *International Journal of Advanced Robotic Systems*, 2013, vol. 10 (87): 1–11 <https://www.researchgate.net/publication/258225996_Improving_Language_Models_in_Speech-Based_Human-Machine_Interaction>.
- Karpov, Alexey, and Markov, Konstantin, and Kipyatkova, Irina, and Vazhenina, Daria, and Ronzhin, Andrey. "Large vocabulary Russian speech recognition using syntactico-statistical language modeling." *Speech Communication*, 2013, Vol. 56: 213–228.
- Kipyatkova, Irina and Karpov, Alexey. "Study of Morphological Factors of Factored Language Models for Russian ASR." *Speech And Computer*. Eds. Ronzhin, Andrey et al. Switzerland: Springer, 2014, 451–458.
- Kipyatkova, Irina, and Karpov, Alexey. "Recurrent Neural Network-based Language Modeling for an Automatic Russian Speech Recognition System." *Proceedings of AINL-ISMW FRUCT Conference*. Eds. Balandin, Sergey et al. St. Petersburg, 2015, 33–38.
- Laptev, Aleksandr, and Korostik, Roman, and Svishev, Aleksey, and Andrusenko, Andrei, and Medennikov, Ivan, and Rybin, Sergey. *You Do Not Need More Data: Improving End-To-End Speech Recognition by Text-To-Speech Data Augmentation*, 2020 <<https://arxiv.org/abs/2005.07157v2>>.
- Mikolov, Tomas et al. *Distributed Representations of Words and Phrases and their Compositionality*, 2013 <<https://arxiv.org/abs/1310.4546v1>>.
- Mikolov, Tomas et al. *Efficient Estimation of Word Representations in Vector Space*, 2013 <<https://arxiv.org/abs/1301.3781v3>>.
- O'Shaughnessy, Douglas. "Invited paper: Automatic speech recognition: History, methods and challenges." *Pattern Recognition*, 2008, 41: 2966–2967 <<https://www.sciencedirect.com/science/article/abs/pii/S0031320308001799>>.
- Raffel, Collin, and Shazeer, Noam, and Roberts, Adam, and Lee, Katherine, and Narang, Sharan, and Matena, Michael, and Zhou, Yanqi, and Li, Wei, and Liu, Peter J. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." *Journal of Machine Learning Research*, 2020, Vol. 21: 1–67.
- Schneider, Steffen, and Baevski, Alexei, and Collobert, Ronan, and Auli, Michael. *wav2vec: Unsupervised Pre-training for Speech Recognition*, 2019 <<https://arxiv.org/abs/1904.05862v4>>.

- Shavrina, Tatiana, and Shapovalova, Olga. "To the Methodology of Corpus Construction for Machine Learning: 'Taiga' Syntax Tree Corpus and Parser." *Proceedings of the International Conference „Corpus Linguistics–2017”*. Zakharov, Viktor Pavlovich, Khokhlova, Mariya Vladimirovna (eds.). St. Petersburg, 2017, 78–84.
- Tampel', Ivan Borisovich, and Karpov, Aleksey Anatol'yevich. *Avtomaticheskoye raspoznavaniye rechi. Uchebnoye posobiye*. Sankt-Peterburg: Universitet ITMO, 2017 [Тампель, Иван Борисович, and Карпов, Алексей Анатольевич. *Автоматическое распознавание речи. Учебное пособие*. Санкт-Петербург: Университет ИТМО, 2017].
- Tampel', Ivan Borisovich. "Avtomaticheskoye raspoznavaniye rechi — osnovnyye etapy 50 za let." *Nauchno-tekhnicheskyy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki*, 2015, Vol. 15, no. 6: 957–968 [Тампель, Иван Борисович. "Автоматическое распознавание речи — основные этапы за 50 лет." *Научно-технический вестник информационных технологий, механики и оптики*, 2015, Vol. 15, no. 6: 957–968].
- Vaswani, Ashish (et al.). *Attention Is All You Need*, 2017 <<https://arxiv.org/abs/1706.03762v5>>.
- Wolf, Thomas (et al.). *Transformers: State-of-the-Art Natural Language Processing*, 2020 <<https://aclanthology.org/2020.emnlp-demos.6.pdf>>.
- Yakovenko, Olga, and Bondarenko, Ivan, and Borovikova, Mariya, and Vodolazsky, Daniil. "Algorithms for automatic accentuation and transcription of russian texts in speech recognition systems." *Speech And Computer*. Eds. Karpov, Alexey, and Jokisch, Oliver, and Potapova, Rodmonga. Switzerland: Springer, 2018, 768–777.
- Ziółko, Bartosz, and Ziółko, Mariusz. *Przetwarzanie mowy*. Kraków: Wydawnictwa AGH, 2011.



ANDRZEJ CHARCIAREK

 ORCID: <https://orcid.org/0000-0001-5651-5730>

Uniwersytet Śląski w Katowicach

KORPUS RÓWNOLEGŁY A EKWIWALENTYZACJA

PARALLEL CORPUS AND EQUIVALENCING

The article is devoted to possibilities of using the InterCorp Polish-Russian parallel corpus in the process of searching for translations equivalents. It also presents resources and functionalities of the InterCorp, and concentrates on how it can be used by Russian philologists. It is increasingly useful in linguistic and translational research. A particular focus is put on the process of equivalencing i.e. searching for and establishing equivalents and a significant role of the parallel corpus which allows for finding appropriate translation solutions that sometimes can pose a problem for a single translator or lexicographer. The article also presents a deeper reflection on the significance and prospects of empirical research in the contemporary linguistics.

Badań na gruncie językoznawstwa korpusowego prowadzonych w polskim środowisku rusycystycznym nie znajdziemy wiele. Najlepszym tego dowodem jest monografia *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*¹, stanowiąca w zamierzeniu jej redaktora, Wojciecha Chlebdy, „obudowę przysłownikową *Podręcznego idiomatykonu polsko-rosyjskiego*”², a poświęcona zastosowaniu korpusów w badaniach lingwistycznych. Wśród autorów poszczególnych rozdziałów znaleźli się zarówno twórcy korpusów, jak i ich użytkownicy: Milena Hebal-Jezierska, Piotr Fliciński, Piotr Pęzik, Tadeusz Piotrowski, Łukasz Grabowski, Mirosław Bańko, Marek Iwanowski, Piotr Wierchoń i Jan Wawrzyńczyk. Już sama analiza składu autorów dowodzi, że rusycyści są

¹ W. Chlebda (red.), *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*, Wydawnictwo Uniwersytetu Opolskiego, Opole 2013.

² W. Chlebda, *Korpusologia użytkowa dla początkujących i zaawansowanych*, w: W. Chlebda (red.), *Na tropach translatów...*, s. 7–15.

reprezentowani jedynie przez Jana Wawrzyńczyka (redaktora monografii, rusycystę, pomijam). I choć od wydania monografii minęło wiele lat, to niewiele się zmieniło. W dalszym ciągu badania korpusowe szerzej i intensywniej rozwijane są w polskim środowisku anglistycznym, polonistycznym czy bohemistycznym. Na taki stan rzeczy składa się wiele czynników, jak choćby tworzenie korpusów narodowych i zaangażowanie w nie przedstawicieli poszczególnych filologii³. Nie sposób również nie dostrzegać znaczenia tradycji badań korpusowych, które powstały i rozwijały się głównie w środowisku anglistycznym⁴.

W niniejszym tekście, choć będzie on głównie poświęcony roli korpusu równoległego InterCorp⁵ w procesie ekwiwalentyzacji⁶, przywołam na wstępie jeszcze dwa korpusy, które — jak zakładam — są najbardziej znane w środowisku rusycystycznym.

Pierwszy z nich to Polsko-Rosyjski i Rosyjsko-Polski Korpus Równoległy Uniwersytetu Warszawskiego⁷ zawierający 30 milionów segmentów. Należy dodać, że za segmenty uważa się nie tylko tradycyjnie rozumiane formy wyrazowe, lecz także aglutynacyjne morfemy osobowo-liczbowe czasownika. Na jego zawartość składają się teksty literackie, prasowe, religijne i prawne. Jego zaletą z pewnością są przekłady bezpośrednie polsko-rosyjskie i rosyjsko-polskie dodatkowo uzupełnione polskimi i rosyjskimi przekładami tekstów literackich w języku angielskim, niemieckim i francuskim. Proporcje ułożyły się następująco: 50% oryginałów polskich, 33% oryginałów rosyjskich i 17% oryginałów z języków trzecich. Można różnie oceniać proporcje poszczególnych tekstów, zwłaszcza 90-procentowy udział tekstów literackich może wydawać się dla jednych użytkowników zaletą, dla drugich — wadą. Ważną informacją jest ograniczony dostęp do 34% tekstów, które są dostępne tylko dla współpracowników korpusu. Z pozostałych 66% tekstów można korzystać bez logowania.

³ Należy tu wspomnieć także o projekcie *Wielkiego słownika języka polskiego PAN* powstającym głównie w oparciu o teksty Narodowego Korpusu Języka Polskiego.

⁴ Nie przypadkiem pierwszy polski podręcznik językoznawstwa korpusowego powstał w środowisku anglistycznym: B. Lewandowska-Tomaszczyk (red.), *Podstawy językoznawstwa korpusowego*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2005.

⁵ <https://intercorp.korpus.cz>.

⁶ To właśnie ten termin znalazł się w tytule, ponieważ — moim zdaniem — bardzo dobrze oddaje specyfikę pracy z korpusem równoległym.

⁷ <http://pol-ros.polon.uw.edu.pl>.

Drugi, polsko-rosyjski korpus równoległy, stanowiący integralną część Narodowego korpusu języka rosyjskiego⁸, liczy ponad 6 milionów słów. Wśród tekstów znajdziemy sporo polskiej i rosyjskiej literatury klasycznej (Sienkiewicz, Prus, Briusow, Sołogub), ale także nielicznych autorów współczesnych (Sapkowski, Chmielewska). Zawartość polsko-rosyjskiego korpusu równoległego uzupełniają teksty prawne i artykuły prasowe. Niestety niewielkie zasoby tego korpusu praktycznie uniemożliwiają pozyskiwanie odpowiedniej liczby poświadczeń do analizy korpusowej.

Trzeci, to wspomniany polsko-rosyjski korpus równoległy InterCorp stanowiący integralną część Narodowego korpusu języka czeskiego (Český národní korpus), regularnie wzbogacany o nowe teksty. InterCorp jest korpusem wielojęzycznym (ostatnia wersja 13. gromadzi 40 języków), w którym język czeski zajmuje pozycję uprzywilejowaną. Nie sposób nie wspomnieć o mocnej pozycji języka angielskiego, który poza czeskim jest drugim językiem interfejsu. Każdy czeski tekst w korpusie posiada odpowiednik, którym może być obcojęzyczny oryginał lub jego tłumaczenie na język obcy. Najpełniejszy opis InterCorp znajdziemy w *Praktycznym przewodniku po korpusie równoległym InterCorp*⁹ Elżbiety Kaczmarskiej i Alexandra Rosena.

Każdy rok przynosi nową wersję korpusów poszczególnych języków, których zasoby są systematycznie powiększane. W momencie powstawania niniejszego tekstu użytkownicy mogą korzystać z 13. wersji udostępnionej w 2020 roku. W wersji tej wielkość korpusu rosyjskiego po raz pierwszy przekroczyła liczbę 20 milionów słów, a polskiego — zbliżyła się do 90 milionów słów. Dowodem na sukcesywnie rosnące rozmiary korpusu polskiego i rosyjskiego jest porównanie ich ostatnich trzech wersji (wielkości podane w liczbach bezwzględnych):

- polski moduł InterCorp: 11 wersja — 86 230 000, 12 wersja — 87 567 000, 13 wersja — 88 161 000;
- rosyjski moduł InterCorp: 11 wersja — 18 282 000, 12 wersja — 19 560 000, 13 wersja — 20 078 000.

Analiza zasobów pokazuje, że oba korpusy — polski i rosyjski — powiększają swoje rozmiary dzięki włączeniu do nich tekstów beletrystycznych. Wielkość poszczególnych kolekcji z przyczyn często

⁸ <https://ruscorp.ru>.

⁹ E. Kaczmarska, A. Rosen, *Praktyczny przewodnik po korpusie równoległym InterCorp*, w: *Praktyczny przewodnik po korpusach języków słowiańskich*. M. Hebal-Jeziarska (red), Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa, s. 207–231.

obiektywnych pozostaje praktycznie niezmienna. Zjawisko to dotyczy kolekcji, które nie mogą być powiększane, ponieważ tworzą zamknięte zbiory tekstów. Przykładem tego jest kolekcja tekstów publicystycznych z nieistniejącego już serwisu www.presseurop.eu. Analogiczna sytuacja ma miejsce w przypadku zbioru zawierającego zapisy obrad Parlamentu Europejskiego z lat 2007–2011 z korpusu Europarl.

Co zrozumiałe, korzystanie z poszczególnych kolekcji zależy od tego, czy są obecne w obu zestawianych ze sobą językach. To one tworzą korpus równoległy tekstów sobie odpowiadających, mogący skutecznie posłużyć do pracy translatorskiej czy leksykograficznej.

O tym, jak istotne znaczenie ma wielkość i powiększanie zasobów korpusowych i, co się z tym bezpośrednio wiąże, przydatność korpusu do wspomnianych badań, świadczą liczby poświadczeń tych samych wyrażeń¹⁰ w poszczególnych jego wersjach, na przykład:

wyrażenie	InterCorp	wersja 13	Wersja 12	Wersja 11
<i>złożyć podanie</i>	pol.-ros.	12	11	– ¹¹
	pol.-czes.	55	54	2
<i>spokojna głowa</i>	pol.-ros.	12	12	12
	pol.-czes.	81	81	79
<i>niedźwiedzia przysługa</i>	pol.-ros.	1	1	–
	pol.-czes.	20	19	18 ¹²
<i>dyrektor finansowy</i>	pol.-ros.	50	50	50
	pol.-czes.	59	59	59
<i>rodzona matka</i>	pol.-ros.	2	2	5
	pol.-czes.	17	16	29
<i>walić głową w mur</i>	pol.-ros.	4	4	3
	pol.-czes.	12	12	12

Jak widać, choć każda nowa wersja InterCorp jest większa od poprzedniej, to wyszukiwanie określonych wyrażeń w najnowszej wersji nie oznacza automatycznie większej liczby poświadczeń. Dobrą ilustracją tego zjawiska jest wyrażenie *rodzona matka*, które w polsko-czeskiej 11. wersji posiada znacznie więcej wyników (29), niż w naj-

¹⁰ Choć są wyjątki.

¹¹ Wyrażenie *złożyć podanie* nie ma poświadczeń w wyszukiwaniu z atrybutem *lemat*, natomiast za pomocą atrybutu *forma wyrazowa (word)* uzyskujemy 4 poświadczenia.

¹² W wersji 11. InterCorp wyszukiwanie z wykorzystaniem atrybutu *lemat* ciągu *niedźwiedzi przysługa* nie przynosi żadnego poświadczenia. Uzyskane 18 wyników to rezultat wyszukiwania *niedźwiedź przysługa*. Najpewniej przyczyną takiego stanu rzeczy był błąd w anotacji, który już w kolejnej 12. wersji został usunięty.

nowszej 13. (17). Fakt ten może skłaniać użytkownika do korzystania ze starszych wersji InterCorp, tym bardziej że taka możliwość istnieje, ponieważ wszystkie wersje, począwszy od 6., są wciąż dostępne. Znaczenie wielkości korpusu dobrze obrazuje liczba poświadczeń lematu *książka* czy *wziąć*, jednostek ogólnoodmianowych, w poszczególnych wersjach korpusu:

wyraz	InterCorp	wersja 13	wersja 12	wersja 11
<i>książka</i>	pol.-ros.	3380	3129	2788
	pol.-czes.	16299	16201	15836
<i>wziąć</i>	pol.-ros.	7111	6882	6513
	pol.-czes.	41404	41128	39554

Uzyskany materiał empiryczny pozwala na sformułowanie kilku uwag natury ogólniejszej.

Po pierwsze, im większy korpus, tym więcej poświadczeń. Zjawisko to widać zarówno między zestawianymi językami (polsko-rosyjski, polsko-czeski), jak i ich kolejnymi wersjami polsko-rosyjskiego czy polsko-czeskiego InterCorp.

Po drugie, skala wzrostu korpusu polsko-rosyjskiego nie dotyczy — co znamienne — w równym stopniu wszystkich notowanych w nim jednostek. Można wykazać, że dodanie napisów filmowych spowoduje wzrost liczby rejestrowanych w nim jednostek potocznych, dodanie tekstów administracyjnych — typowych dla nich sformułowań.

Po trzecie, korpus InterCorp nie zwiększa swoich rozmiarów w sposób proporcjonalny; między poszczególnymi wersjami mogą być mniejsze lub większe różnice. Przeprowadzona sonda odzwierciedla w zasadzie proporcje między przywołanymi wyżej wielkościami omawianych korpusów — większy wzrost można zaobserwować między wersją 11. a 12. niż między 12. a 13.

Niekorzystnym zjawiskiem jest to, że oprócz wiązanych ręcznie tekstów beletrystycznych, tworzących rdzeń (czes. jądro), zasobami, mogącymi służyć do ustanowienia ekwiwalencji polsko-rosyjskiej pozostają tylko dwie kolekcje wiązane automatycznie: napisy filmowe z amatorskiego serwisu OpenSubtitles oraz przekłady Biblii. Trzecia kolekcja tekstów rosyjskich dostępna w InterCorp — teksty prasowe z platformy Project Syndicate, nie może być wykorzystana, ponieważ brakuje jej polskich odpowiedników. Korpus polsko-rosyjski dysponuje więc tekstami ograniczonymi pod kątem ich zróżnicowania gatunkowego; wyraźnie brakuje tekstów prawnych, publicystycz-

nych czy ekonomicznych. O tym, jaki ma to wpływ na wyszukiwanie w korpusie, można się przekonać, porównując polsko-czeski InterCorp z polsko-rosyjskim. Kwerenda najbardziej typowych terminów prawnych i ekonomicznych przynosi następujące rezultaty (liczby poświadczeń):

wyrażenie	Polsko-rosyjski Intercorp	Polsko-czeski InterCorp
<i>przedmiot umowy</i>	0	145
<i>umowa o pracę</i>	2	439
<i>zabrania się</i>	3	211
<i>decyzja w sprawie</i>	0	2582
<i>wzrost gospodarczy</i>	0	2621
<i>obligacje państwowe</i>	2	99
<i>lokata terminowa</i>	0	6
<i>niniejszym informować</i>	0	13
<i>zapytanie ofertowe</i>	1	2
<i>rynek międzybankowy</i>	0	21

Powyższe wartości dobrze dokumentują wagę obecności tekstów prawnych i ekonomicznych w polsko-czeskim InterCorp, dzięki którym pozyskiwane są większe lub mniejsze liczby poświadczeń. Jednocześnie brak tego typu tekstów w korpusie polsko-rosyjskim uniemożliwia praktycznie pozyskanie jakichkolwiek wyników. W praktyce oznacza to, że użytkownik nie jest nawet naprowadzany na ewentualny ekwiwalent czy ekwiwalenty. Źródło pochodzenia poszczególnych poświadczeń można określić za pomocą opcji *typy tekstów* (czes. *typy textů*):

wyrażenie	Korpus	Liczba poświad- czeń	Core	Acquis	Euro- parl	Press- Europ	Sub- titles
<i>przedmiot umowy</i>	pol.ros.	–	–	–	–	–	–
	pol.-czes.	145	10	127	8	–	–
<i>umowa o pracę</i>	pol.ros.	2	2	–	–	–	–
	pol.-czes.	439	15	298	114	10	3
<i>zabrania się</i>	pol.-ros.	3	1	–	–	–	2
	pol.-czes.	211	10	154	16	7	24
<i>decyzja w sprawie</i>	pol.ros.	–	–	–	–	–	–
	pol.-czes.	2582	18	1811	693	50	10
<i>wzrost gospodarczy</i>	pol.ros.	–	–	–	–	–	–
	pol.-czes.	2621	1	540	1794	280	6

KORPUS RÓWNOLEGŁY A EKWIWALENTYZACJA

<i>obligacje państwowe</i>	pol.-ros.	2	2	–	–	–	–
	pol.-czes.	99	3	39	12	43	2
<i>lokata terminowa</i>	pol.-ros.	–	–	–	–	–	–
	pol.-czes.	6	–	5	–	1	–
<i>niniejszym informować</i>	pol.-ros.	1	1	–	–	–	–
	pol.-czes.	13	2	8	2	–	1
<i>zapytanie ofertowe</i>	pol.-ros.	–	–	–	–	–	–
	pol.-czes.	2	1	1	–	–	–
<i>rynek międzybankowy</i>	pol.-ros.	–	–	–	–	–	–
	pol.-czes.	21	–	8	10	3	–

Jak ważna jest weryfikacja pozyskanego materiału przekonuje rejestracja przez wyszukiwarkę Kontext poświadczeń, mających odmienną semantykę. Wprowadzenie zapytania *zabraniać się* z atrybutem *lemat* pozwala na uzyskanie wyników w znakomitej większości tego zwrotu, który był przewidywany, czyli *zabrania się* (od czasownika *zabraniać się*). Jednakowoż zwrot ten, typowy dla tekstów administracyjnych typu rozporządzenie czy zarządzenie, przybiera taką samą postać, jak *zabrania się*, będący formą rzeczownika odsłownego *zabranie się* w dopełniaczu, np. *dasz mi szansę zabrania się na samolot*. Przykład ten (z części polsko-czeskiej InterCorp) dowodzi obowiązku sprawdzenia pozyskanych poświadczeń, ponieważ domysł czy intuicja użytkownika mogą okazać się zawodne.

Na korpus polsko-rosyjski, zresztą na pozostałe w InterCorp również, składają się głównie tłumaczenia, czyli przekład bezpośredni polsko-rosyjski lub rosyjsko-polski ma miejsce rzadko.

Właśnie InterCorp ze względu na swoją unikatowość skłania do głębszej refleksji nad miejscem korpusu równoległego w badaniach lingwistycznych, a szczególnie w procesie ustalania ekwiwalentów przekładowych. I to z dwu przyczyn.

Pierwsza jest oczywista. Wykorzystanie korpusów równoległych, zresztą nie tylko tego rodzaju, w rusycystycznych badaniach językoznawczych jest niewielkie. Dobrze więc byłoby ten stan rzeczy zmienić.

Druga przyczyna, niejako wynikająca z pierwszej, to wskazanie jego zalet odnośnie do pracy translatorskiej i leksykograficznej, nie mówiąc o jego przydatności w badaniach glottodydaktycznych czy socjolingwistycznych. O każdym z wymienionych pól zastosowań można napisać odrębnie i obszernie zarazem. Tu skupimy się wyłącznie na procesie poszukiwania ekwiwalentów przekładowych, czyli na procesie, który Wojciech Chlebda nazywa ekwiwalentyzacją. Jego

zdaniem, przyjęcie perspektywy patrzenia na pracę tłumacza czy leksykografa przekładowego wymusza dodanie do pojęć *ekwiwalent* (na gruncie leksykografii — jednostki języka) i *ekwiwalencja* (relacja wzajemnej odpowiedniości między jednostkami) pojęcia trzeciego, wspomnianego już pojęcia ekwiwalentyzacji, rozumianej jako proces poszukiwania i dochodzenia do jednostek pozostających w danej relacji¹³. Teksty korpusowe są chyba jednym z najlepszych narzędzi, aby ten proces obserwować. To właśnie korpusy równoległe najlepiej obrazują to, co Chlebda nazywa „stopniowym osaczeniem odpowiednika przekładowego”¹⁴. W rzeczy samej korpus równoległy może stanowić szczególnego rodzaju „słownik propozycji translatorskich”¹⁵ ze wszystkimi słabościami i ograniczeniami, ale też niewątpliwymi zaletami. Wybory translatorskie znajdujące się w profesjonalnych pracach leksykograficznych są najczęściej autorstwa jednego czy kilku-kilkunastu specjalistów leksykografów. Propozycje w korpusie równoległym natomiast to rozwiązania heterogeniczne — przedstawiciele różnych generacji, grup społecznych i zawodowych, w różnym wieku i różnej płci. W odróżnieniu od indywidualnego idiolektu konkretnego słownikarza czy tłumacza, w korpusie równoległym znajdujemy setki czy tysiące poświadczeń, rejestrujących idiolekty ich autorów. Wszystko to stwarza zupełnie nową sytuację, w której możliwe jest wskazanie takich odpowiedniości przekładowych, które przekraczają możliwości jednego użytkownika języka, nawet biorąc pod uwagę jego ponadstandardowe zdolności intelektualne i językowe. To ograniczenie dotyczy bez wyjątku każdego użytkownika, który korzysta „z takich możliwości, jakie oferuje mu jego język, ale w ramach tych możliwości sam wybiera sposób obrazowania”¹⁶.

Wielkość i zróżnicowanie tekstów korpusu są niezbędne dla skutecznej analizy statystycznej i, co z tym się bezpośrednio wiąże, ekwiwalentyzacji. W przypadku polsko-rosyjskiego InterCorp — jak już stwierdziłem i zilustrowałem na konkretnych przykładach — o dużych możliwościach nie może być mowy, ponieważ i jedna, i druga wartość nie jest duża. Rzecz jasna, to wcale nie przesądza o jego nieprzydatności w procesie tłumaczenia określonych typów tekstów.

¹³ W. Chlebda, *Ekwiwalencja i ekwiwalenty: między słownikiem a tekstami*, w: W. Chlebda (red.), *Na tropach tłumaczeń...*, s. 41.

¹⁴ Tamże, s. 42.

¹⁵ Tamże.

¹⁶ E. Tabakowska, *Gramatyka i obrazowanie. Wprowadzenie do językoznawstwa kognitywnego*, Wydawnictwo Oddziału PAN, Kraków 1995, s. 57.

Wszystko zależy od tego, jakie jednostki mają być tłumaczone. Zasoby korpusowe, czyli gatunki tekstów zgromadzonych w korpusie, decydują o jego przydatności w wyszukiwaniu ekwiwalentów określonych translandów. Nieobecność w korpusie tekstów prawnych czy ekonomicznych automatycznie spowoduje, że użytkownik nie znajdzie w nim np. skonwencjonalizowanych sformułowań typowych dla decyzji administracyjnej czy rozporządzenia. Brak tego typu gatunków administracyjnych cechujących się własnymi sposobami realizacji tekstowych oraz wyspecjalizowanymi formami językowymi skąpe praktycznie każdą kwerendę korpusową na niepowodzenie. Wydaje się, że wiedza ta, choć oczywista, nie jest powszechna. Niewielka jest świadomość faktu, że w przypadku korpusu równoległego stworzenie odpowiednio dużego korpusu jest prawdziwym wyzwaniem; pozyskiwanie tekstów napotyka różnego rodzaju trudności, choćby wynikające z ochrony praw autorskich. Porównanie zasobów korpusu polsko-rosyjskiego i polsko-czeskiego InterCorp dobrze zjawisko to ilustruje. Jeśli pierwszy zawiera teksty beletrystyczne i dwie kolekcje (Biblia, napisy filmowe z platformy OpenSubtitles), to drugi dysponuje, oprócz literatury pięknej, pięcioma kolekcjami: aktami prawnymi Unii Europejskiej z korpusu *Acquis Communautaire*, zapisami posiedzeń Parlamentu Europejskiego z lat 2007–2011, tekstami publicystycznymi z serwisu www.presseurop.eu, napisami filmowymi z platformy OpenSubtitles i Biblią.

Aby uzmysłowić sobie znaczenie wielkości korpusu i różnorodności tekstów w nim zgromadzonych, można porównać liczbę poświadczeń wybranych jednostek wielowyrazowych w układzie polski — rosyjski i polski — czeski¹⁷:

wyrażenie	InterCorp polsko-rosyjski	InterCorp polsko-czeski
<i>złożyć podanie</i>	12 poświadczeń	55 poświadczeń
<i>spokojna głowa</i>	12 poświadczeń	88 poświadczeń
<i>niedźwiedzia przysługa</i>	1 poświadczenie	20 poświadczeń
<i>dyrektor finansowy</i>	50 poświadczeń	59 poświadczeń
<i>rodzona matka</i>	2 poświadczenia	17 poświadczeń
<i>walić głowę w mur</i>	4 poświadczenia	12 poświadczeń

Już wstępna analiza danych korpusowych pokazuje, że w wielu przypadkach przeprowadzenie analizy korpusowej i wskazanie par

¹⁷ Wyniki dla wyszukiwania z atrybutem *lemat*.

ekwiwalentów nie jest możliwe; poświadczeń jest zbyt mało. Widać dużą różnicę w liczbie pozyskanych poświadczeń między korpusem równoległym polsko-czeskim a polsko-rosyjskim na korzyść tego pierwszego. Jednocześnie zestawienie poświadczeń dokumentuje nie tylko znaczenie wielkości korpusu, ale także zróżnicowanie gatunkowe zgromadzonych w nim tekstów.

Dla pozyskania jak największej liczby poświadczeń wykorzystałem atrybut *lemat (lemma)*, który pozwala uwzględnić wszystkie formy utworzone od wpisanych leksemów *walić głowa w mur*, np. *walisz głową w mur*, *walą głową w mur*¹⁸.

O tym, jak istotne jest zbudowanie zapytania i weryfikacja automatycznie pozyskanych danych korpusowych, przekonuje poszukiwanie rosyjskich i czeskich ekwiwalentów polskiego wyrażenia *spokojna głowa*. W *Wielkim słowniku języka polskiego* IJP PAN jest ono definiowane następująco: ‘*pot.* wypowiedziane w celu uspokojenia osoby, która się czymś martwi’¹⁹. Pozostawienie atrybutu *lemat* i wpisanie w wyszukiwarkę wyrazów hasłowych *spokojny głowa* umożliwia pozyskanie 12 par polsko-rosyjskich i 88 polsko-czeskich. Niemniej po bardziej szczegółowym przyjrzeniu się pozyskanym poświadczeniom okazuje się, że wśród zgromadzonego materiału znajdują się takie, o które w analizie nie chodziło:

Potem już ze spokojną głową będę mógł znaleźć jakieś odpowiednie zajęcie, które by mi dało przyzwoity zarobek i pozostawiło dosyć wolnego czasu na pisanie;

Wyjdziesz ze spokojną głową;

A ja upodobałem sobie ciebie i tę młodą kobietę, toteż powiedziałem tylko słówko i możecie już mieć spokojną głowę.

Jak widać, przytoczone poświadczenia (*ze spokojną głową, mieć spokojną głowę*), rejestrują inne znaczenie, nienotowane w słownikach, a które można ująć jako ‘nie zaprzętać sobie głowy innymi problemami’.

W przypadku badania wyrażenia *spokojna głowa*, występującego wyłącznie w tej jednej utrwalonej postaci, lepszym rozwiązaniem jest zatem użycie atrybutu *forma wyrazowa (word)*, który spowoduje wyszukanie jedynie wyrażenia *spokojna głowa* (81 poświadczeń).

Inną strategię wyszukiwawczą należałoby zastosować wobec wyrażenia *niedźwiedzia przysługa*, które może być w tekście poprzedzone

¹⁸ Kontext wyszukuje także przykłady z rzeczownikiem odczasownikowym, tj. *walenie głową w mur*.

¹⁹ <https://wsjp.pl/haslo/podglad/18949/spokojna-glowa>.

czasownikami *wyświadczyć, wyrządzić, oddać czy zrobić*²⁰. Użycie atrybutu *lemat* zapewnia maksymalne wykorzystanie zasobów korpusowych, choć w przypadku korpusu rosyjskiego niewiele to zmienia (1 poświadczenie). Jeszcze raz wielkość tego korpusu uniemożliwia jakąkolwiek analizę korpusową.

Dodatkowo pośród pozyskanych poświadczeń mogą się znajdować (i się znajdują) pary przekładowe, które takimi nie są. Wskazywane przez Kontext, podstawową wyszukiwarkę InterCorp, translaty nie mają niekiedy żadnego związku z translandem. Przyczyną takiego stanu rzeczy najczęściej jest automatyczne wiązanie, które zestawia ze sobą zdania nawzajem sobie nieodpowiadające. To zjawisko naturalne i poniekąd nieuniknione, z którym użytkownik InterCorp powinien się oswoić. Jest ono mniej istotne niż mała liczba poświadczeń, będąca rezultatem przeszukiwania korpusu niewielkich rozmiarów.

W tym miejscu należy odnieść się szczegółowiej do zagadnienia, jakim jest właściwe sformułowanie zapytania, które w praktyce oznacza optymalne wyszukiwanie materiału. Znaczenie prawidłowego zbudowania zapytania trudno przecenić. Wielokroć analiza korpusowa okazuje się nieskuteczna z powodu niewłaściwego określenia badanych jednostek i ich cech gramatycznych. Potencjalne błędy mogą być zatem następstwem przyjęcia błędnych atrybutów dla wyjściowej jednostki – translandu. Zwraca na to uwagę Milena Hebal-Jezińska w tekście *Podstawowe zasady korzystania z korpusów przy badaniu języka*²¹, w którym to znajdziemy wiele istotnych wskazówek dotyczących analizy korpusowej. Właśnie prawidłowe zbudowanie zapytania może mieć wpływ na liczbę pozyskanych poświadczeń. Ma to szczególne znaczenie w przypadku korpusu InterCorp, który dysponuje bogatą anotacją wewnętrzną i zewnętrzną²². W pracy z korpusem można skorzystać z analizy wieloczynnikowej, na przykład zawężając zakres wyszukiwania ekwiwalentów do określonej kolekcji. W przypadku polsko-rosyjskiego InterCorp, choć wyboru praktycznie nie ma (poza tekstami beletrystycznymi można skorzystać tylko

²⁰ Elektroniczny *Wielki słownik języka polskiego IJP PAN* wskazuje czasowniki *wyświadczyć, oddać i zrobić*.

²¹ M. Hebal-Jezińska, *Podstawowe zasady korzystania z korpusów przy badaniu języka*, w: W. Chlebda (red.), *Na tropach translatów...*, s. 17–30.

²² E. Kaczmarek, *Metody ustalania ekwiwalentów czasownikowych wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2019, s. 44–46.

z tekstu Biblii i napisów filmowych), można ograniczyć na przykład wyszukiwanie wyrażen potocznych wyłącznie do kolekcji napisów filmowych. Ponadto w każdym wyszukiwaniu użytkownik ma do dyspozycji wiele atrybutów, które mogą zawęzić pole badawcze do ściśle wyselekcjonowanych tekstów. Bogata anotacja zewnętrzna umożliwia określenie podstawowych cech tekstu istotnych dla ustanowienia takich ekwiwalencji jak: przynależność do określonej części korpusu (beletrystyka [core] lub kolekcje [text.group]), typ tekstu (text.txttype), autor (text.autor), tłumacz (text.translator), miejsca wydania (text.pubplace), rok wydania (text.pubyear), tytuł tekstu (text.title), język oryginału (text.srclang), płeć autora (text.authsex), płeć tłumacza (text.transsex). InterCorp rozstrzyga kwestię zasadniczą dla badań nad ekwiwalencją, jaką jest określenie, czy tekst jest oryginałem, czy przekładem (text.original). W zależności od charakteru badań dzięki atrybutom można zawęzić pole wyszukiwawcze wyłącznie na przykład do układu tekst oryginalny – tekst przekładu. W przypadku zaawansowanych i systematycznych badań możliwe jest stworzenie własnego subkorpusu, ograniczającego przeprowadzane kwerendy do określonego typu tekstów, na przykład tekstów beletrystycznych czy administracyjnych²³. Można na przykład stworzyć subkorpus złożony z tekstów oryginalnych i ich przekładów bezpośrednich na określony język. Zbudowanie subkorpusu z wykorzystaniem określonych atrybutów jest rozwiązaniem korzystnym ze względu na uniknięcie ewentualnych błędów w ciągłym wprowadzaniu filtrów i sformułowania w konsekwencji pochopnych wniosków. Wystarczy przywołać przytoczone wyżej wyrażenie *spokojna głowa*, aby się w tym utwierdzić. Utworzenie bowiem zapytania *spokojna głowa* z zastosowaniem atrybutu *lemat* w korpusie polsko-rosyjskim nie dostarczy żadnych poświadczeń. Natomiast wyszukiwanie z użyciem wyrazów hasłowych *spokojny głowa* pozwoli uzyskać 12 poświadczeń. To wciąż mało, tym bardziej że wśród ewentualnych, sugerowanych przez InterCorp, ekwiwalentów znajdują się wyrażenia będące rezultatem błędnego automatycznego wiązania. Mimo to badacz pozyskuje materiał wyjściowy do dalszej analizy. Podobne błędy użytkowników w budowaniu zapytania przewidzieli zresztą twórcy InterCorp, usta-

²³ Sposób utworzenia subkorpusu znajdziemy w: E. Kaczmarek, A. Rosen, *Praktyczny przewodnik po Korpusie Równoległym InterCorp*, w: *Praktyczny przewodnik po korpusach języków słowiańskich*, M. Hebal-Jezińska (red), Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2014, s. 219.

nawiając od wersji 13. atrybut *lemma / word* (*lemat / forma wyrazowa*) jako domyślny.

Niemniej należy mieć na uwadze fakt, że uwzględnianie poszczególnych filtrów automatycznie pociąga za sobą zmniejszenie zakresu przeszukiwanych zasobów tekstowych, co w przypadku korpusów niewielkich rozmiarów, a do takich należy polsko-rosyjski InterCorp, może znacznie ograniczać liczbę pozyskiwanych poświadczeń, a w skrajnym przypadku uniemożliwić kwerendę.

Trudno nie odnotować w tym miejscu kluczowej roli użytkownika, któremu w świetle sformułowanych wyżej czynności z pewnością niełatwo sprostać sporym wyzwaniom. Te ostatnie wymagają bowiem od niego nie tylko odpowiednich kompetencji językowych, interkulturowych i komunikacyjnych, ale także umiejętnego wykorzystania możliwości, jakie stwarza korpus równoległy.

Aby wskazać przydatność korpusu równoległego w ustalaniu wspomnianej już ekwiwalencji przekładowej, niezbędne jest zdefiniowanie tej ostatniej, ponieważ jest ona terminem kluczowym. *Ekwiwalencja* jako termin doczekała się licznych opracowań teoretycznych we współczesnym przekładoznawstwie²⁴. Obecność korpusu równoległego w procesie ustalania par przekładowych powoduje, że najbliższe jest mi traktowanie ekwiwalencji jako konstytutywnej właściwości przekładu posiadającej charakter gradualny. Właśnie stopniowalność (zmienność) ekwiwalencji jako jej cecha podstawowa najbardziej wyeksponowana jest w opracowaniach Wojciecha Chlebdy i Romana Lewickiego. Pierwszy z nich określa typy ekwiwalentów, drugi natomiast cechy ekwiwalencji przekładowej.

Chlebda wyróżnia cztery typy ekwiwalentów:

1) ekwiwalenty pełne (adekwatne) — są ekwiwalentne wobec jednostki wyjściowej pod względem sensu i obrazowości, parametrów pragmatycznych, stylowych, a często również gramatycznych,

2) ekwiwalenty funkcjonalne — są ekwiwalentne wobec jednostki wyjściowej pod względem sensu i funkcji komunikacyjnej, różnią się obrazowością i strukturą formalną (składniową, gramatyczną),

3) ekwiwalenty zachwiane — struktury treściowe ekwiwalentów pozostają ekwiwalentne tylko częściowo wobec jednostek wyjściowych,

²⁴ M.in. J.C. Catford, *A Linguistic Theory of Translation. An Essay in Applied Linguistics*, Oxford University Press, Oxford 1965; V.N. Komissarov, *Теория перевода (лингвистические аспекты)*, Москва 1990; R. Lewicki, *Zagadnienia lingwistyki przekładu*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 2017; W. Chlebda, *Ekwiwalencja i ekwiwalenty...*

4) ekwiwalenty zerowe – w języku tłumaczenia nie odnotowuje się gotowego sposobu wyrażania danej treści pochodzącej z języka oryginału²⁵.

Chyba najbardziej istotne z naszej perspektywy jest traktowanie powyższych typów ekwiwalentów jako ekwiwalentów słownikowych. Powyższa klasyfikacja uzmysławia asymetrię systemów pojęciowo-leksykalnych, w tym przypadku polszczyzny i ruszczyzny²⁶.

Lewicki wskazuje sześć cech ekwiwalencji przekładowej:

1) Ma charakter asymetryczny (jednokierunkowy). Jeden z tekstów jest oryginałem, drugi – jego przekładem. Odwrócenie tego układu nie jest możliwe.

2) Jest cechą przekładu i polega na podobieństwie przekładu do oryginału poprzez jego naśladowanie.

3) Jest stopniowalna, co oznacza, że przekład jest stopniowalny – może być w mniejszym lub większym stopniu ekwiwalentny wobec oryginału.

4) Jest względna, ponieważ może być oceniana wedle różnych kryteriów.

5) Jest hierarchiczna, ponieważ podlega jej nie tylko przekład jako całość, ale także jego poszczególne części.

6) Nie jest oparta na kryterium równoznaczności wyrażen językowych użytych w oryginale i w jego przekładzie. Jej efekt stanowi uzyskanie przekładu równoważnego w stosunku do oryginału. Zatem zarówno oryginał, jak i przekład powinny posiadać analogiczną wartość komunikacyjną dla ich odbiorców²⁷.

Choć obaj badacze koncentrują się na relacji między oryginałem a przekładem i ich wykładnikami, to jednak można dostrzec u nich dwie odmienne perspektywy. Chlebda reprezentuje podejście leksykograficzne, nastawione na notowanie par przekładowych jako jednostek języka w słowniku z całą złożonością tego procesu. U Lewickiego bardziej widać podejście tekstowe – nastawione na proces tworzenia ekwiwalentnego tekstu zarówno pod względem jego całości, jak i tworzących go elementów.

Uwzględniając powyższe, można uznać, że poszczególne przekłady sytuują się na kontinuum od najmniej do najbardziej ekwi-

²⁵ W. Chlebda, *Ekwiwalencja i ekwiwalenty...*, s. 38–40.

²⁶ Przytaczane przez autora przykłady dotyczą leksykografii przekładowej polsko-rosyjskiej.

²⁷ R. Lewicki, *Zagadnienia lingwistyki przekładu*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 2017, s. 138–139.

walentnego wobec określonego oryginału. Określenia „najmniej ekwiwalentny” czy „najbardziej ekwiwalentny” to, oczywiście, pojęcia względne. Stopień ekwiwalencji – jak słusznie zauważa Lewicki – może być określany według różnych kryteriów. W literaturze przedmiotu najczęściej wskazuje się kryteria semantyczne, formalne, pragmatyczne i formalne. W zależności od autora to czy inne kryterium jest najczęściej uprzywilejowane, co często zależne jest od zakresu ekwiwalencji zachodzącej bądź między fragmentami tekstów (np. jednostkami wielowyrazowymi), bądź między całymi tekstami.

Kryterium, które przyjmuję tu za podstawę ustanawiania ekwiwalencji przekładowej, a w którym korpus równoległy może odgrywać ważną rolę, to kryterium sytuacji użycia. Właśnie ona stanowi podstawy porównania – *tertium comparationis* (dalej: TerComp). Pojęcie to w ostatnich latach na dobre zadomowiło się w badaniach nad ekwiwalencją przekładową²⁸. Świadomie pomijam tu opracowania na gruncie innych nauk czy dyscyplin, w których *tertium comparationis* służy również jako podstawa porównania. Wypada w tym miejscu przywołać jego opis pierwotny, jeszcze z lat 70. autorstwa Andrzeja Bogusławskiego, twierdzącego, że

rozstrzygającą podstawą identycznościową, czyli ich ogólnym i najwyższym *tertium comparationis* jest konkretna sytuacja wypowiedzi, znaczy to, że staramy się utworzyć taki tekst w języku przekładu, który normalny nosiciel tego języka utworzyłby w sposób maksymalnie naturalny w sytuacji właściwej tekstowi wyjściowemu²⁹.

Do zagadnienia TerComp Bogusławski powraca w monografii *Podstawy konfrontatywnej lingwistyki przekładowej*, uznając kryterium sytuacyjności za najważniejsze i zwracając uwagę na atomistyczny charakter konfrontacji przekładowej³⁰. Zdaniem badacza, nie istnieją żadne ograniczenia dotyczące wyboru określonego translatu, w szczególności ze strony języka wyjściowego, czyli języka translan-

²⁸ Zob. M. Mocarz-Kleindienst, *Tertium comparationis w badaniach nad przekładem filmowym*, „Rocznik Przekładoznawczy”, 2020, s. 233–248; W. Chlebda, *Ekwiwalencja i ekwiwalenty...*, s. 21–44; T. Piotrowski, *Tertium comparationis w przekładoznawstwie*, w: P. Stalmaszczyk (red.), *Metodologie językoznawstwa. Od genu języka do dyskursu*, Łódź, s. 175–191.

²⁹ A. Bogusławski, *Problem tertium comparationis w porównaniu lingwistycznym*, „Kwartalnik Neofilologiczny” 1976, XXIII(3), s. 295–303.

³⁰ A. Bogusławski, *Podstawy konfrontatywnej lingwistyki przekładowej*, Oficyna Wydawnicza Leksem, Łask 2013, s. 79–83.

du. Jedyne ograniczenia wskazania translatu nakłada sytuacja użycia translandu.

Analogicznie wspomnianą sytuację wypowiedzi przedstawia Piotrowski:

Zacznijmy od opisu typowej sytuacji tłumaczenia, w której występuje indywidualny akt wypowiadania. Ów akt niesie sąd, należy więc od niego wyjść. Analizujemy wobec tego sytuację wypowiadania zdania w L2, po czym ma nastąpić jego tłumaczenie na zdanie L1. Patrząc na taką sytuację czysto zewnątrznie, można wyróżnić kilka jej składowych: użytkownika języka L2, użytkownika języka L1, tłumacza między oboma, oraz kontekst, w jakim się znaleźli. Częściami tej sytuacji, kontekstu, są formy znakowe (semiotyczne), ze szczególnym uwzględnieniem form języka. Do form znakowych zaliczamy jednak także znaki niewerbalne, np. ubiór i inne oznaki statusu społecznego uczestników (tłumacz jest zwykle niejako neutralny), a także wszelki inny kontekst historyczno-społeczny, jak wiedzę o tym, co doprowadziło do zaistniałej sytuacji i co z niej może wyniknąć (np. względy dyplomatyczne, transakcje handlowe, potrzeby turysty, etc.)³¹.

Ponadto autor podkreśla, że tłumacz w całym procesie tłumaczenia powinien zachować postawę neutralną, nie wygłaszać sądów i nie przejawiać emocji. Co istotne, Piotrowski powyższe uwagi o TerComp odnosi nie tylko do procesu tłumaczenia tekstu mówionego, ale także tekstu pisanego.

W przypadku ustalania ekwiwalencji przekładowej z użyciem korpusu równoległego pojęcie TerComp nabiera specyficznego znaczenia. Jeśli przyjąć, że InterCorp zawiera w większości przekłady, to jako TerComp należałoby wskazać sytuację, w której powstaje oryginał (często angielski). To właśnie sytuacja użycia angielskiego translandu wskazuje właściwe ekwiwalenty polskie, rosyjskie czy czeskie. Mamy zatem sytuację bardziej złożoną niż ma to miejsce w klasycznym układzie tłumaczenia bezpośredniego transland — translat. Relacja naśladowania translandu przez translat nie ma miejsca; mamy natomiast zestawienie dwu translatów w większym czy mniejszym stopniu ekwiwalentnych wobec obcojęzycznego translandu. Choć zjawisko to jest niepożądane, ponieważ uobecnienie angielszczyzny czy innego języka może w pewien sposób zaburzyć prowadzoną analizę, to jednak przy dużej liczbie poświadczeń nie ma zasadniczego wpływu na wskazywane ekwiwalenty. Wskazywane zatem przez InterCorp pary przekładowe to w znakomitej większości zestawienia translatów. Deficyt tłumaczeń bezpośrednich, w których nie ujawnia się obecność angielszczyzny (głównie) jako języka pośrednika, nie jest z pewnością zaletą InterCorp.

³¹ T. Piotrowski, *Tertium comparationis w przekładoznawstwie. Metodologie językoznawstwa. Od genu języka do dyskursu*, Stalmaszczyk, Piotr (red.), Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2011, s. 175–191

Jednocześnie trzeba wyraźnie zaznaczyć, że stworzenie odpowiedniej wielkości korpusu równoległego polsko-rosyjskiego czy polsko-czeskiego, zawierającego wyłącznie tłumaczenia bezpośrednie, jest praktycznie niewykonalne. Bezspornym walorem InterCorp jest uwzględnianie tłumaczeń zarówno profesjonalnych (np. literatura piękna), jak i amatorskich (np. napisy filmowe). W odniesieniu do zasobów polsko-rosyjskiego InterCorp można mówić właśnie o tych dwu sytuacjach.

Ogląd sytuacji w tekście literackim jest przecież odmienny niż w przypadku sporządzania nieprofesjonalnych napisów filmowych. W pierwszej czynność tę wykonuje profesjonalny tłumacz, w drugiej – tłumacz amator. Zróżnicowanie to z pewnością pomaga użytkownikowi korpusu podjąć trafne decyzje ekwiwalentyzacyjne.

Jak wiadomo, InterCorp zawiera teksty pisane, w zależności od zestawianych ze sobą języków mniej lub bardziej zróżnicowane gatunkowo i tematycznie. W przypadku układu polski – rosyjski wspomniane zróżnicowanie jest niewielkie. Nie licząc tekstu Biblii, wprowadzonego w wersji 10. z 2017 roku, do dyspozycji użytkownika korpusu pozostaje wspomniana literatura piękna oraz amatorskie napisy filmowe.

Wśród beletrystyki znajdziemy zarówno literaturę klasyczną, jak i współczesną. Zróżnicowanie zestawianych ze sobą przekładów jest spore. Kontext zestawia przekłady na polski i rosyjski nie tylko z angielskiego, ale także z innych języków, np. serbskiego (np. Ivo Andrić), czeskiego (np. Jaroslav Hašek, Milan Kundera), szwedzkiego (np. Astrid Lindgren), włoskiego (np. Umberto Eco), portugalskiego (np. Paulo Coelho) czy fińskiego (np. Mika Waltari). Zestawiane mogą być także polskie lub rosyjskie oryginały z przekładami na te języki (Fiodor Dostojewski, Michaił Bułhakow, Vladimir Nabokov, Boris Pasternak, Czyngiz Ajtmatow, Aleksander Puszkina, Nikołaj Ostrowskij, Ilja Ilf i Jewgienij Pietrow, Olga Tokarczuk, Stanisław Lem). Jak widać, bezpośrednie przekłady z polskiego na rosyjski czy odwrotnie stanowią jedynie uzupełnienie przekładów na polski czy rosyjski z innych języków.

Kolekcją istotną dla poszukiwania polsko-rosyjskich par przekładowych są przywołane już amatorskie napisy filmowe z platformy OpenSubtitles, która jest serwisem społecznościowym. Nie chcę tu szczegółowo omawiać cech napisów sporządzanych przez amatorów, ponieważ zrobiłem to już w artykule *Napisy filmowe jako źródło ekwiwalentów przekładowych w polsko-czeskim korpusie równo-*

*ległym InterCorp*³². Zastąpienie układu języków z polsko-czeskiego na polsko-rosyjski w praktyce niewiele zmienia. Autorzy najczęściej opracowują napisy w podobnych warunkach – bez specjalistycznego zaplecza technicznego. Kryteria czasowo-przestrzenne tworzenia napisów nie są przez nich ściśle przestrzegane. Niekiedy ma się wręcz wrażenie, że nie wszyscy tłumacze są świadomi ich istnienia. Ich wiedza językowa, dotycząca zarówno języka źródłowego (zwykle angielski), jak i przekładu (polski, rosyjski), jest zróżnicowana. Nie są więc zaskoczeniem liczne błędy językowe (gramatyczne, leksykalne, stylistyczne) czy tłumaczeniowe. Dla przykładu można przywołać typowe ortograficzne błędy w poświadczeniach polskich i rosyjskich:

<i>агентство</i> (zamiast <i>агентство</i>)	20 poświadczeń
<i>мороженное</i> (zamiast <i>мороженое</i>)	5 poświadczeń
<i>вообщем</i> (zamiast <i>в общем</i>)	16 poświadczeń
<i>wziąść</i> (zamiast <i>wziąć</i>)	30 poświadczeń
<i>napewno</i> (zamiast <i>na pewno</i>)	76 poświadczeń
<i>rozumie</i> (zamiast <i>rozumiem</i>)	5 poświadczeń

Jeśli jednak porównamy liczbę przytoczonych błędów z liczbą form prawidłowych:

<i>агентство</i>	68 poświadczeń
<i>мороженное</i>	118 poświadczeń
<i>в общем</i>	342 poświadczenia
<i>wziąć</i>	509 poświadczeń
<i>na pewno</i>	1396 poświadczeń
<i>rozumiem</i>	1723 poświadczenia

to okaże się, że poświadczenia zawierające błędy, nawet te najbardziej typowe i rozpowszechnione, stanowią zdecydowaną mniejszość.

Wynika z tego, że do pozyskanych poświadczeń należy podchodzić z dużą choć nieprzesadną ostrożnością. Obraz ekwiwalencji polsko-rosyjskiej w przypadku napisów filmowych może być dodatkowo zniekształcony przez błędy, będące rezultatem automatycznego wiązania, bo takie zostało zastosowane. Niemniej błędy te pozostają w zdecydowanej mniejszości, tworząc margines zestawianych nieade-

³² A. Charciarek, *Napisy filmowe jako źródło ekwiwalentów przekładowych w polsko-czeskim korpusie równoległym InterCorp*, „Studia Rossica Posnaniensia” 2021, nr 1, s. 39–49.

kwatnych par przekładowych. Zalet wykorzystania napisów w ustanawianiu ekwiwalencji określonych jednostek językowych jest znacznie więcej. Dzięki tłumaczeniom sporządzanym przez autorów posiadających różne cechy społeczno-demograficzne (płeć, wykształcenie, wiek, środowisko) pozyskiwane tłumaczenia, a następnie powstające z nich pary przekładowe bywają w znacznym stopniu zróżnicowane. Nie oznacza to wcale, że zróżnicowanie to jest zjawiskiem niepożądanym; wręcz przeciwnie, dobrze obrazuje to, co rozumiemy pod pojęciem *wariantywność przekładu*. Korpus równoległy jest w tym względzie narzędziem nieocenionym, rejestrującym wybory ekwiwalentyzacyjne tłumaczy i jednocześnie demonstrującym poszukiwania optymalnego przekładu. W procesie tym — jak się wydaje — tłumacz pozostaje elementem kluczowym, dokonującym kolejnych wyborów mniej lub bardziej udanych w dużym stopniu zdeterminowanych jego wiedzą (nie tylko językową), doświadczeniem czy predyspozycjami. Jeśli uznamy — a wydaje się to oczywiste — że tekst wyjściowy może posiadać więcej niż jeden przekład, to potwierdzenie tego znajdziemy właśnie w korpusie równoległym. W niektórych przypadkach uzyskujemy wyniki idące w tysiące, niekiedy w setki, a czasami w dziesiątki poświadczeń. To pozwala na stworzenie klasyfikacji ewentualnych translatów według kryterium statystycznego, które może przesądzić o wyborze określonego ekwiwalentu.

Szczególne wartości napisy filmowe mają w odniesieniu do jednostek występujących w odmianie mówionej języka, których brak w korpusie równoległym jest szczególnie odczuwalny. Decydują o takim stanie rzeczy względy obiektywne. Jak wiadomo, teksty mówione zwykle nie są tłumaczone, a jeśli już, to rzadko. Mamy więc do czynienia poniekąd z deficytem systemowym tekstów mówionych w korpusie równoległym, który w pewien sposób napisy filmowe rekompensują.

Celem niniejszego tekstu nie było zademonstrowanie procesu ustalenia pary przekładowej za pomocą narzędzi i zasobów korpusowych, czyli wyjściem od określonych translatów i zakończeniem w postaci wskazania ich możliwych translatów. Ten ostatni etap został przeze mnie świadomie pominięty, ponieważ sam w sobie może stanowić odrębne opracowanie. Skupiłem się tu na etapie pośrednim, tj. sytuacji, gdy mamy już wyodrębnione translady i staramy się określić zakres możliwych propozycji translatorskich, spośród których wytypowane będą translaty. Na etapie tym kluczową rolę może odegrać korpus równoległy. Co oczywiste, po spełnieniu określonych warunków. Wśród nich znajdują się te, które zależą tak od korpusu (wielkość oraz zróżnicowanie tematyczne i gatunkowe jego zasobów,

udostępnione funkcje wyszukiwawcze), jak i użytkownika (umiejętne posługiwanie się narzędziami korpusowymi, wiedza i kompetencja odnośnie do zestawianych języków). Pamiętać jednak należy, że nawet w przypadku spełnienia tych warunków użytkownik korpusu jest zobowiązany do zweryfikowania pozyskanego materiału.

Choć w porównaniu z polsko-czeskim polsko-rosyjski InterCorp korpus równoległy stwarza mniejsze możliwości w ustanawianiu par przekładowych, to może okazać się on przydatny we wskazywaniu ekwiwalentów określonych jednostek, zwłaszcza tych, które występują w języku ogólnym (beletrystyka) i odmianie potocznej (napisy filmowe). Niestety, przeprowadzenie ekwiwalentyzacji w odniesieniu do słownictwa specjalistycznego nie jest możliwe; brak tekstów o tematyce prawniczej, ekonomicznej, medycznej itp. praktycznie wyklucza pozyskanie jakichkolwiek poświadczeń.

Podsumowując, można stwierdzić, że InterCorp to narzędzie, które może skutecznie posłużyć do tłumaczenia określonych fragmentów tekstu i wskazywania możliwych ekwiwalentów. Jednocześnie korzystanie z InterCorp wymaga od użytkownika wiedzy i doświadczenia w posługiwaniu się funkcjami, które może on wykorzystywać.

Można z dużym prawdopodobieństwem założyć, że wraz z powiększaniem rozmiarów i doskonaleniem funkcji wyszukiwawczych polsko-rosyjski InterCorp będzie zyskiwał coraz więcej na znaczeniu w badaniach nad przekładem czy leksykografią przekładową. W ten sposób pozwoli on na dostarczanie danych empirycznych dotyczących istnienia bądź nieistnienia określonych jednostek języka, a w przypadku istnienia — danych statystycznych (frekwencyjnych), podnosząc rangę badań ilościowych i łącząc je z badaniami jakościowymi. Rysuje się zatem niezwykle obiecująca perspektywa dla współczesnych badań leksykograficznych i przekładoznawczych oraz językoznawczych w ogóle.

REFERENCES

- Bogusławski, Andrzej. *Podstawy konfrontatywnej lingwistyki przekładowej*. Łask: Oficyna Wydawnicza Leksem, 2013.
- Bogusławski, Andrzej. "Problem tertium comparationis w porównaniu lingwistycznym." *Kwartalnik Neofilologiczny*, 1976, z. 3: 295–303.
- Catford, John Cunnison. *A Linguistic Theory of Translation. An Essay in Applied Linguistics*. Oxford: Oxford University Press, 1965.
- Charciarek, Andrzej. "Napisy filmowe jako źródło ekwiwalentów przekładowych w polsko-czeskim korpusie równoległym InterCorp." *Studia Rossica Posnaniensia*, 2021, vol. 46, no. 1: 39–49.

KORPUS RÓWNOLEGŁY A EKWIWALENTYZACJA

- Chlebda, Wojciech (ed.). *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2013.
- Chlebda, Wojciech. "Ekwiwalencja i ekwiwalenty: między słownikiem a tekstami." *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Ed. Chlebda, Wojciech. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2011: 21–44.
- Chlebda, Wojciech. "Korpusologia użytkowa dla początkujących i zaawansowanych." *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Ed. Chlebda, Wojciech. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2013: 7–15.
- Hebal-Jeziarska, Milena. „Podstawowe zasady korzystania z korpusów przy badaniu języka.” *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Ed. Chlebda, Wojciech, Opole: Wydawnictwo Uniwersytetu Opolskiego, 2013: 17–30.
- Kaczmarek, Elżbieta, Rosen, Alexander. „Praktyczny przewodnik po korpusie równoległym InterCorp.” *Praktyczny przewodnik po korpusach języków słowiańskich*. Ed. Hebal-Jeziarska, Milena. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2014: 207–231.
- Kaczmarek, Elżbieta. *Metody ustalania ekwiwalentów czasownikowych wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2019.
- Komissarov, Vilen Naumovich. *Teoriya perevoda (lingvisticheskiye aspekty)*. Moskwa: Vysshaya shkola, 1990 [Комиссаров, Вилем Наумович. *Теория перевода (лингвистические аспекты)*]. Moskwa: Высшая школа, 1990].
- Lewandowska-Tomaszczyk, Barbara. Ed. *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2005.
- Lewicki, Roman. *Zagadnienia lingwistyki przekładu*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 2017.
- Moczarz-Kleindienst, Maria. "Tertium comparationis w badaniach nad przekładem filmowym." *Rocznik Przekładoznawczy*, 2020: 233–248.
- Piotrowski, Tadeusz. "Tertium comparationis w przekładoznawstwie." *Metodologie językoznawstwa. Od genu języka do dyskursu*. Ed. Stalmaszczyk, Piotr. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2011: 175–191.
- Tabakowska, Elżbieta. *Gramatyka i obrazowanie. Wprowadzenie do językoznawstwa kognitywnego*. Kraków: Wydawnictwo Oddziału PAN, 1995.

ONLINE SOURCES

- Bańczyk, Łucja, Dybalska, Renata, Vavřín, Martin. Korpus InterCorp — polština, verze 13 z 1. 11. 2020. Ústav Českého národního korpusu FF UK, Praha 2020 <<http://www.korpus.cz>> (10.12.2021).
- Rajnochová, Natálie, Runštuková, Naděžda, Vavřín, Martin. Korpus InterCorp — ruština, verze 13 z 1. 11. 2020. Ústav Českého národního korpusu FF UK, Praha 2020 <<http://www.korpus.cz>> (10.12.2021).
- Rosen, Alexandr, Vavřín, Martin, Zasina, Adrian Jan. Korpus InterCorp — čeština, verze 13 z 1. 11. 2020. Ústav Českého národního korpusu FF UK, Praha 2020 <<http://www.korpus.cz>> (10.12.2021).
- <<https://ruscorporpora.ru>>.
- <<http://pol-ros.polon.uw.edu.pl>>.



DARIA DAYTER

 <https://orcid.org/0000-0001-5054-5330>

Tampere University

ŁUKASZ GRABOWSKI

 <https://orcid.org/0000-0002-3968-9218>

Uniwersytet Opolski

CORPUS-BASED AND CORPUS-DRIVEN RESEARCH ON TRANSLATION AND INTERPRETING IN RUSSIAN: THE PAST, THE PRESENT AND THE FUTURE

CORPUS-BASED AND CORPUS-DRIVEN RESEARCH ON TRANSLATION AND INTERPRETING IN RUSSIAN:
THE PAST, THE PRESENT AND THE FUTURE

Recently, there has been a growing interest in descriptive and applied studies on translation conducted with the use of methodologies grounded in corpus linguistics. In this paper, we present an overview of state-of-the-art research in corpus-based and corpus-driven translation and interpreting studies conducted with the use of Russian corpora, notably parallel and comparable ones. In contrast to research conducted on other European languages (English, French, German, Spanish etc.), the considerable research tradition in Russian remains relatively unknown to a wider scholarly audience. We outline the scope of research conducted so far, present the most important parallel and comparable corpora with a Russian subcorpus used in translation and interpreting studies, discuss the state-of-the-art research methods, including descriptive and inferential statistics, and summarize selected studies of considerable impact in the discipline. Finally, we identify research gaps and outline avenues for future research.

1. INTRODUCTION

The claim put forward almost thirty years ago by Albrecht Neubert and Gregory M. Shreve¹ that contemporary translation studies represent a cluster of overlapping disciplines and that there is no single way or approach to study translational phenomena is even more pertinent nowadays. One of the research strands that have gained mo-

¹ A. Neubert, G. Shreve, *Translation as Text*, The Kent State University Press, Kent-London 1992, p. 6.

mentum since the 1990s and early 2000s, i.e. since the publication of classic works by James Holmes², Mona Baker³, Miriam Shlesinger⁴, Maria Tymoczko⁵, Sara Laviosa⁶ and Maeve Olohan⁷, among others, is descriptive studies on translation and interpreting conducted with the use of corpus linguistic methodologies (corpus-based and corpus-driven ones). This research direction offers new ways of conceptualizing and studying translation and interpreting largely from, but not limited to, a quantitative, statistical perspective,⁸ also in an attempt to introduce more methodological rigour and objectivity. According to Olohan⁹, corpus linguistic research methods allowed, among others, to reveal what is probable and typical in translation and combine quantitative and qualitative research methods in linguistic descriptions encompassing lexical, syntactic and discoursal features of translational and non-translational texts.

Three main research orientations in corpus linguistic descriptive translation studies have become particularly popular in the last three decades or so. These are research on translation/interpreting universals, translator's/interpreter's style and translation/interpretation style.¹⁰ As for translation universals, they constitute generalizations based on repeatedly observed textual characteristics (lexical, gram-

² J. Holmes, *The Name and Nature of Translation Studies*, in: Lawrence Venuti (Ed.), *The Translation Studies Reader*, Routledge, London 1988/2000, pp. 172–185.

³ M. Baker, *Corpus Linguistic and Translation Studies: Implications and applications*, in: M. Baker, G. Francis, and E. Tognini-Bonelli (Eds), *Text and Technology, in Honour of John Sinclair*. John Benjamins, Amsterdam 1993, pp. 233–250; M. Baker, *Corpora in Translation Studies: An Overview and Some Suggestions for Future Research*, “Target” 1995, no. 7(2), pp. 223–243.

⁴ M. Shlesinger, *Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies*, “Meta” 1998, no. 43(4), pp. 86–493.

⁵ M. Tymoczko, *Computerized Corpora and the Future of Translation Studies*, “Translators’ Journal” 1998, 43(4), pp. 652–660.

⁶ S. Laviosa, *Corpus-based translation studies: theory, findings, applications*. Rodopi, Amsterdam 2002.

⁷ M. Olohan, *Introducing Corpora in Translation Studies*. Routledge, London 2004.

⁸ This reflects the claim made by Holmes (1988/2000) who argues that descriptive Translation Studies should aim to describe, predict and explain translational phenomena using authentic empirical data with a goal of producing a systematic description of a translation product (text), process and function.

⁹ M. Olohan, *Introducing Corpora in Translation Studies...*

¹⁰ These research orientations mentioned in this paragraph are discussed in greater detail in Grabowski (2015).

matical, discoursal etc.) typical of translated texts irrespective of particular pairs of source and target languages¹¹, while the translator's style represents a personal attribute conceptualized as an individual and consistently used set of features of translation (or interpreting) that it shares with other translations (or interpretations) completed by one and the same translator or interpreter, irrespective of source texts.¹² Finally, translation style has been seen as a textual attribute rather than a personal one and it refers to the description of the way a translator or interpreter responded to a source text.¹³ In the last two approaches, the style is largely conceptualized as "a property of texts constituted by an ensemble of formal features which can be observed quantitatively and qualitatively" and its analysis may involve "any linguistic feature that can be formally defined and measured computationally."¹⁴

Recently, corpus-based translation and interpreting studies have enjoyed considerable and growing popularity.¹⁵ A perfunctory inspection of the most recent studies in the field (Kotze¹⁶, Dayter¹⁷; Delaere, and De Sutter¹⁸; De Sutter, and Lefer¹⁹; Kajzer-Wietrzny,

¹¹ M. Baker, *Corpus Linguistic and Translation Studies...*

¹² J. Rybicki, *Stylometric translator attribution: do translators leave lexical traces?*, T. Piotrowski, Ł. Grabowski (Eds), *The Translator and the Computer*. Wydawnictwo Wyższej Szkoły Filologicznej, Wrocław 2013, pp. 193–204.

¹³ K. Malmkjær, *What Happened to God and the Angels: An Exercise in Translational Stylistics*. "Target" 2003, no. 15(1), pp. 37–58; G. Saldanha, *Translator style. Methodological considerations*, "The Translator" 2011, 17(1), pp. 25–50.

¹⁴ J. B. Herrmann, K. van Dalen-Oskam, and Ch. Schöch, *Revisiting Style, a Key Concept in Literary Studies*, "Journal of Literary Theory" 2015, pp. 43–44.

¹⁵ M.-A. Lefer, *Parallel Corpora*. In: M. Paquot, and S. Gries (Eds), *A Practical Handbook of Corpus Linguistics*, Springer Nature Switzerland AG, Cham 2020, pp. 258.

¹⁶ H. Kotze, *Converging what and how to find out why: An outlook on empirical translation studies*, L. Vandevoorde, J. Daems, and B. Defrancq (Eds), *New Empirical Perspectives on Translation and Interpreting*. Routledge, Abingdon 2020, pp. 333–370.

¹⁷ D. Dayter, *Describing lexical patterns...*; D. Dayter, *Strategies in a corpus of simultaneous interpreting. Effects of directionality, phraseological complexity, and position in speech event*. "Meta: Translator's Journal" 2021, no. 66(3), pp. 594–617; D. Dayter, *Collocations in non-interpreted and simultaneously interpreted English: a corpus study*, In: L. Vandevoorde, J. Daems, and B. Defrancq (Eds), *New Empirical Perspectives on Translation and Interpreting*, Routledge, London 2020, pp. 67–91.

¹⁸ I. Delaere, and G. De Sutter, *Applying a Multidimensional, Register-Sensitive Approach to Visualize Normalization in Translated and Non-Translated Dutch*, "Belgian Journal of Linguistics" 2013, no. 25, pp. 43–60.

¹⁹ G. De Sutter, and M.-A. Lefer, *On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach*, "Perspectives" 2020, no. 28(1), pp. 1–23.

and Grabowski²⁰) attests to the growing awareness among researchers, notably those specializing in variationist linguistics, of the limitations of basic quantitative (descriptive statistics) and qualitative (concordance analysis) research methods. These have turned out to be insufficient to address more complex problems of linguistic variation in translation and interpreting, and researchers instead resort to more sophisticated statistical methods (multivariate statistics etc.). This point has been elaborated recently by Stefan Th Gries²¹, Gert DeSutter and Marie-Aude Lefer²², and Maria Calzada Pérez and Sara Laviosa²³: the latter two argue that monofactorial (i.e. based on a single independent explanatory variable) descriptive studies conducted by corpus linguists, e.g. related to frequency differences of a certain linguistic feature (dependent variable) between translational and non-translational texts (independent explanatory variable), do not contribute much to our understanding of translation or interpreting. In fact, there are a number of other factors that impact the use of language in a particular way in translation or interpreting and that is why it is necessary to use more sophisticated statistical methods²⁴ (e.g. regression models) to account for multifactorial empirical phenomena²⁵. Thus, identification and comparisons of frequencies of individual linguistic features or pairs of source-text items and their translation equivalents in texts or corpora do not offer a plausible explanation of the observed phenomenon. In other words, putting

²⁰ M. Kajzer-Wietrzny, and Ł. Grabowski, *Formulaicity in constrained communication: An intermodal approach*, “MonTI” 2021, no. 13, pp. 148–183.

²¹ S. Gries, *The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models*, *Corpora* 2015, no. 10(1), pp. 95–125; S. Gries, *On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally*, “Journal of Second Language Studies” 2018, no. 1(2), pp. 276–308.

²² G. De Sutter, and M.-A. Lefer, *On the need for a new research agenda...*

²³ M. Calzada Pérez, and S. Laviosa, *Twenty-five years on: Time to pause for a new agenda for CTIS*. In: M. Calzada Pérez, and S. Laviosa (Eds), *Reflexión crítica en los estudios de traducción basados en corpus / CTS spring-cleaning: A critical reflection*, “MonTI” 2021, no. 13, p. 20.

²⁴ For more on statistical modeling of linguistic data (linear models, generalized linear models and mixed models etc.), see Winter (2019).

²⁵ This constitutes an interesting counterpoint to the developments in corpus linguistics, where there have been recent calls for backgrounding sophisticated statistical reporting in favour of more traditional linguistic analysis (Egbert, Larsson, and Biber 2020; Larsson, Egbert, Biber forthcoming). Also, a critical reflection on quantitative approaches to linguistic research, replication crisis etc. can be found in Sönning and Werner (2019).

forward an explanation without the use of multifactorial statistics is tantamount to mere speculation and should be discouraged in corpus linguistic research on translation and/or interpreting.

In what follows, we will present and discuss the scope of state-of-the-art corpus linguistic research on translation and interpreting conducted with the use of Russian language material (and set it against the backdrop of the recent advances in the field). We will also present the most important parallel and comparable corpora²⁶ of Russian compiled so far and used by researchers interested in descriptive translation and interpreting studies. We believe that while the work in the main European scholarly languages such as French and German has been recognised and well-integrated in the research landscape of translation and interpreting studies, e.g. Marianne Lederer²⁷ or Silvia Kalina²⁸, the considerable research tradition in Russian remains largely unknown.

2. CORPUS-BASED TRANSLATION STUDIES: AN OVERVIEW OF CORPORA WITH RUSSIAN LANGUAGE

Russian linguistics has developed from a strong philological tradition that comprises linguistic and literary studies under one umbrella, often locating them in the same research department. As a result, linguists have often worked on literary data to make their research relevant and applicable in this broader disciplinary context. This situation has led to proliferation of parallel corpora of literary language, and their relative prominence compared to other European languages' resources²⁹.

Among freely available resources, Russian corpus studies boast a Russian-French parallel corpus of poetry³⁰ whose web interface enables both the traditional queries (lemma, wordform) and searching

²⁶ See Zanettin (2012), Mikhailov and Cooper (2016), Ustaszewski and Stauder (2020), among others, for a more detailed discussion of the typology and compilation of various types of corpora used in translation studies research.

²⁷ M. Lederer, *La traduction simultanée. Expérience et théorie*, Lettres modernes Minard, Paris 1981.

²⁸ S. Kalina, *Strategische Prozesse beim Dolmetschen: theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*, Günter Narr, Tübingen 1998.

²⁹ For a more detailed overview of Russian corpora, including monolingual and comparable ones, see Kopotev et al. (2021).

³⁰ A. Balakin, A. Bodrova, S. Orekhov, and M. Rybina. Russian-French Poetry Corpus 1800–1820, 2011, Available at: <http://nevmenandr.net/fr/>.

specifically in the rhyming area. Another freely available parallel corpus of fiction is the Chinese and Russian Literature Corpus (<http://www.rucorpus.cn/>), containing a subcorpus of Anton Chekhov's short stories (Ru>Zh) and a subcorpus of Chinese fiction (Zh>Ru). The literary component of the Polish-Russian Parallel Corpus³¹ incorporates both bidirectional translations Ru<>Pl and translations of literary works from other languages into Russian and Polish. An unusual resource at the intersection of historical and comparative linguistics is the parallel corpus of translations of *The Tale of Igor's Campaign* (*Слово о полку Игореве*), an epic poem in Old Slavic.³² It includes around a hundred intralingual translations into Modern Russian as well as 44 translations into other languages.

All four corpora lend themselves well to interdisciplinary use because they offer intuitive web interfaces. A great number of further parallel corpora of fiction, often compiled for specific research projects, are freely available, but require the researcher to be somewhat technically savvy: they offer a download option rather than a web interface. Among these we could highlight the PELCRA Polish-Russian parallel corpus with 20 Polish and 14 Russian literary works, with 4.25 million words³³; the Chinese-Russian corpora of individual novels, e.g. the hand-aligned corpus for *The Republic of Wine*³⁴ (Piao et al. 2014); HunOr: A Hungarian–Russian Parallel Corpus³⁵, with sam-

³¹ M. Łaziński, M. Kuratczyk, B. Orekhov, and E. Slobodjan, *The Polish-Russian Parallel Corpus and Its Application in the Linguistic Analysis*, "Prace Filologiczne" 2012, vol. LXIII, pp. 209–218; M. Łaziński, and M. Kuratczyk, *Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego*, E. Gruszczyńska, and A. Leńko-Szymańska (Eds), *Polskojęzyczne korpusy równoległe*. Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 2016: 83–96; M. Łaziński, and M. Kuratczyk, *Польско-русский параллельный корпус Варшавского университета*, "Język i Metoda", 2016, no. 3, pp. 85–93].

³² Б. Орехов, *Параллельный корпус переводов «Слова о полку Игореве»: итоги и перспективы*, В. Плулган (Ed.). *Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы*. Санкт-Петербург: Нестор-История, 2009, pp. 462–473.

³³ P. Peżik, M. Ogrodniczuk, and A. Przepiórkowski, *Parallel and spoken corpora in an open repository of Polish language resources*, Human Language Technologies as a Challenge for Computer Science and Linguistics. LTC, Poznań 2011, <http://pelcra.pl/new/polrus> (30.08.2021).

³⁴ Z. Piao, Q. Li, and L. Wang, *Jiyu han-e pingxing yuliaoku de „jiu guo“ dieyinci eyi guilü yanjiu*, "Zhongguo eyu jiaoxue" 2014, no. 3, pp. 46–51.

³⁵ M. Szabo, V. Vincze, and I. Nagy, *HunOr: A Hungarian–Russian Parallel Corpus*, In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA),

ples of literary novels, scientific texts and official documents (with approx. 800,000 tokens in each language); or the Polish-Bulgarian-Russian Parallel Corpus³⁶ including 55 texts (literary novels and legal treaties) translated into the aforementioned languages.

Another type of parallel corpora presents the register of ‘language for special purposes’, e.g. business texts, academic language, political and diplomatic discourse, news articles or marketing texts. These can be found in the PEST corpus with state treaties in Swedish, Russian, Finnish³⁷, in the Corpus of Russian Translations of Social and Political Works, including translations of English and German 18th-century treaties into Russian³⁸, as well as in the UMC 0.1 Czech-Russian-English Multilingual Corpus with internet texts and news articles³⁹. One can also find specialized parallel texts with Russian language data in the constantly growing Open Parallel Corpus collection⁴⁰ that enables one to browse and download parallel text files in various formats (e.g. txt, tmx). Another genre corpus is the Russian Learner Translator Corpus (RuLTC), which comprises translations of newspapers and other English mass-media performed by advanced students of translation. RuLTC contains approx. 2.3 million tokens and has an online search interface freely available to researchers.⁴¹

Finally, a number of large, multi-register corpora are available to the researchers that comprise different text types: fiction, lan-

Istanbul 2012, pp. 2453–2458, http://www.lrec-conf.org/proceedings/lrec2012/pdf/262_Paper.pdf.

³⁶ A. Kisiel, V. Koseska-Toszewa, N. Kotsyba, J. Satoła-Staškowiak, and W. Sosnowski, *Polish-Bulgarian-Russian Parallel Corpus*, CLARIN-PL digital repository, 2016, <http://hdl.handle.net/11321/308>.

³⁷ M. Mikhailov, M. Santalahti, and J. Souma (Eds.), *PEST. A parallel electronic corpus of state treaties*, John Benjamins, Amsterdam 2019.

³⁸ Polskoy, Sergey, and Rjéoutski, Vladislav, *The Corpus of Russian Translations*, 2021. <https://krp.dhi-moskau.org/ru/translations> (27.08.2021).

³⁹ N. Kluyeva, and O. Bojar, *UMC 0.1: Czech-Russian-English Multilingual Corpus*. Proceedings of International Conference Corpus Linguistics, October 2008, pp. 188–195, <https://ufal.mff.cuni.cz/umc/cer/download.php?f=umc-0.1-paper-2008.pdf>.

⁴⁰ J. Tiedemann, *OPUS - Parallel Corpora for Everyone*, “Baltic Journal of Modern Computing” 2016, no. 4 (2): Special Issue: *Proceedings of the 19th Annual Conference of the European Association of Machine Translation (EAMT)*..

⁴¹ A. Kutuzov, and M. Kunilovskaya. *Russian Learner Translator Corpus*, in: P. Sojka, A. Horák, I. Kopeček, K. Pala (Eds.), *Text, Speech and Dialogue*. TSD 2014. Lecture Notes in Computer Science, vol. 8655. Springer, Cham, https://doi.org/10.1007/978-3-319-10816-2_39.

guage for specific purposes, and sometimes even transcriptions of spoken language. These will be well familiar to corpus linguists. There is the Russian National Corpus⁴² (<https://ruscorpora.ru/>, NKRY 2005) of 300 million words covering a breadth of historical periods and genres and offering a large POS-tagged component (manually checked); and ParaSol,⁴³ a project that has similar aims to the Russian National Corpus but is currently at an earlier development stage, and focussing specifically on Slavic languages. Another prominent example of this family is InterCorp 13,⁴⁴ an enormous parallel corpus of 1.551 million words with Czech as a pivot language and 40 further languages, including Russian and many more Slavic languages.

3. CORPUS-BASED INTERPRETING STUDIES: AN OVERVIEW OF RUSSIAN LANGUAGE CORPORA

Corpus-based approaches to interpreting studies have become the new standard, compared to the earlier introspection-based approaches. However, as late as in 2011, Setton still called Corpus-Based Interpreting Studies (CIS) “a cottage industry.” This characteristic referred to the fact that CIS was represented by many disparate studies, comparatively small-scale, rather than a single concerted effort with uniform norms, transcription conventions, and systematic research agenda.

Interestingly, corpus-based interpreting research in Russian goes back several decades. For example, Ghelly Chernov’s groundbreaking research on anticipation in simultaneous interpreting is based on corpus data.⁴⁵ According to the compilers of Chernov’s

⁴² As such, NKRY contains a number of parallel sub-corpora that have been developed since 2005 (Sitchinava 2012).

⁴³ R. von Waldenfels, *Compiling a parallel corpus of slavic languages. Text strategies, tools and the question of lemmatization in alignment*, in: B. Brehmer, V. Zdanova, R. Zimny (Eds), *Beiträge der Europäischen Slavistischen Linguistik: POLYSLAV 9*. Verlag Otto Sagner, München 2006, pp. 123–138.

⁴⁴ F. Čermák, and A. Rosen, *The case of InterCorp, a multilingual parallel corpus*. “International Journal of Corpus Linguistics” 2012, no. 13(3), pp. 411–427.

⁴⁵ Gh. Chernov, *Inference and Anticipation in Simultaneous Interpreting*, Benjamins, Amsterdam 1978/2004.

recently translated edition Setton, and Hild,⁴⁶ Chernov worked with a “shoebox corpus” (a paper-based card index) of the United Nations speeches which he taped and transcribed himself. The tapes were allegedly made available to the publishing house in 1987, but it appears that now both the tapes and the unpublished transcriptions are lost⁴⁷. Chernov’s colleague and co-author Irina Zimnaya also worked with corpora of simultaneous interpreting⁴⁸, although we have no record of this resource. Since much of the publishing effort at the time was pedagogically oriented, we have the resulting textbooks based on the corpus research⁴⁹ but not the corresponding corpora.

Although CIS has been gaining momentum in the last decade, at the moment of writing, there are still very limited corpus resources available for the Russian language. This scarcity can be explained by the fact that most existing interpreting corpora are based on the EUROPARL database, a large parallel corpus extracted from the proceedings of the European Parliament and including 21 European languages. Since the Russian Federation is not a part of the European Union and is therefore not represented in the European Parliament, the resource does not contain a Russian component.

A somewhat similar source of parallel language data including the Russian language is provided by the records of the United Nations’ events and proceedings. Although transcripts of the UN proceedings are not readily available, researchers have access to United Nations Web TV (<http://webtv.un.org/>). The website airs a variety of UN events such as General Assembly, press conferences, press briefings etc., with simultaneous interpretation into the official UN languages in a separate audio channel. The events are aired live and the recordings, including simultaneous interpretation, are stored on the website. The material is in public domain, and the United Nations Publications allows the use of portions of its content for educational and research purposes. While the material needs to

⁴⁶ R. Setton, and A. Hild, *Editors’ critical foreword*, in: Gh. Chernov (Ed.), *Inference and Anticipation in Simultaneous Interpreting*, Benjamins, Amsterdam 2004, pp. IX–XXII.

⁴⁷ Ibidem.

⁴⁸ I. Zimnaya, and Gh. Chernov, *Verojatnostnoe Prognozirovanie v Prozesse Sinkhronnogo Perevoda*. In: A. Leont’ev, N. Zhinkin, and A. Shahnarovich (Eds.), *Predvaritel’nye materialy jeksperimental’nyh issledovanij po psiholingvistike*, IY SSSR, Moscow 1973.

⁴⁹ Eg. A. Shiryaev, *Sinkhronnyj perevod*, Voenizdat, Moscow 1979.

be laboriously transcribed from the video/audio format, this additional step results in cleaner data than the automatically harvested EUROPARL texts.⁵⁰

One existing corpus that includes a Russian component and is based in part on the UN WebTV data is the SIREN.⁵¹ The SIREN is a bidirectional parallel Russian-English corpus of original texts and their simultaneous interpreting into English with circa 235,000 word tokens, transcribed, annotated, POS-tagged and parsed (at the sentence-clause level). The SIREN includes 21 speech events from the WebTV archives, among them, for example, the UN Security Council's meetings or sessions of Universal Periodic Review.⁵² Although the corpus is not freely available in a language depository at the moment of writing, the creator can be contacted to provide the corpus on request.

While the SIREN represents the genre of political discourse, another resource, CoInCoUT (Court Interpreting Corpus at the University of Tampere), focuses on courtroom discourse. CoInCoUT includes Finnish-Russian interpretation of 9 sessions, approx. 49,000 words, with three different interpreters.⁵³

Other existing interpreting corpora with a Russian component are not available to researchers, but are purpose-built for specific studies and described in unpublished PhD dissertations.⁵⁴

⁵⁰ Corpora created from the transcripts of video recordings and the parallel interpreting audio stream are not to be confused with the parallel corpora of the UN written documents, e.g. The United Nations Parallel Corpus 1.0 (Ziemski et al. 2016).

⁵¹ D. Dayter, *Describing lexical patterns...*; D. Dayter, *Collocations in non-interpreted and simultaneously interpreted English: a corpus study*, in: L. Vandevoorde, J. Daems, and B. Defrancq (Eds.), *New Empirical Perspectives on Translation and Interpreting*, Routledge, London 2020, pp. 67–91.

⁵² D. Dayter, *Collocations in non-interpreted and simultaneously interpreted English...*

⁵³ M. Mikhailov, H. Tommola, and N. IsoLahti, *Spoken to Spoken vs. Spoken to Written: Corpus Approach to Exploring Interpreting and Subtitling*, "Polibits" 2010, no. 41, pp. 5–9.

⁵⁴ E.g. A. Gorshkova, *Sopostavitel'noye issledovaniye sposobov dostizheniya ekvivalentnosti v sinkhronnom i pis'mennom perevodakh*. Unpublished Phd dissertation, Rossiyskiy universitet druzhby narodov, Moskva 2004; V. Ilukhin, *Strategii v sinkhronnom perevode*. Unpublished PhD dissertation, Moscow 2001; I. Gurin, *Priyemy rechevoy kompressii prisinkhronnom perevozhde s russkogo yazyka na angliyskiy*. Unpublished Phd dissertation, Moskovskiy gosudarstvennyy universitet, Moskva 2009.

To sum up, while translation corpora are a well-developed segment in Russian corpus linguistics resources, there is still a distinct lack of freely available corpora of interpreting.

4. RESEARCH ON TRANSLATION AND INTERPRETING: THE PAST AND THE PRESENT

A perfunctory inspection of state-of-the-art corpus research on translation and interpreting in Russian shows that the most prominent studies can be thematically grouped into two main areas, namely the studies dealing with identification of specific features of translation and/or interpreting and the studies comparing different translation/interpreting modes.

The first kind of studies typically draw their inspiration from the theories of translation universals⁵⁵ or from the translation/interpreting ‘strategies’.⁵⁶ Both of these attempt to capture universal patterns of linguistic behaviour in translation, which manifest themselves on the level of lexis, syntax, and discourse. For example, early cognitive research focused on the strategy of anticipation in simultaneous interpreting.⁵⁷ Relying on a pre-computer index-card corpus, the researchers described multiple instances of anticipation in French-Russian interpreting and concluded that strategic anticipation is a result of the message redundancy characteristic of all natural languages. One of the resources described in section 3 above, the SIREN corpus, gave rise to comprehensive corpus-based research on universals and strategies in Russian-English simultaneous interpreting. The findings are somewhat contradictory with regard to universals (although one of the language subcorpora conforms to the expectations of translation theory, the other exhibits the opposite trend) and invite further corpus-based investigation of interpreting⁵⁸.

⁵⁵ M. Baker, *Corpora in Translation Studies: An Overview and Some Suggestions for Future Research*, “Target” 1995, no. 7(2), pp. 223–243; A. Chesterman, *Hypothesis about translation universals*, in: G. Hansen, K. Malmkjær, and D. Gile (Eds.), *Claims, Changes and Challenges in Translation Studies*, John Benjamins, Amsterdam 2004, pp. 1–13.

⁵⁶ S. Kalina, *Strategische Prozesse beim Dolmetschen...*

⁵⁷ I. Zimnaya, and Gh. Chernov, *Verojatnostnoe Prognozirovanie v Prozesse Sinkhronnogo Perevoda...*

⁵⁸ D. Dayter, *Describing lexical patterns in simultaneously interpreted discourse...*; D. Dayter, *Strategies in a corpus of simultaneous interpreting...*

The descriptions of linguistic markers that serve to operationalise universals and strategies have been applied in the second thematic area of translation corpus research, the studies comparing various translation and interpreting modes. For example, the work by Lapshinova-Koltunski and the RuLTC team look at translation universals (translationese) that manifest to different degrees e.g. in human vs. machine translation, or performed by novice vs. experienced translators.⁵⁹

A comparison between novice (learner) and professional Russian translations of English mass-media texts with the reference Russian corpus of non-translations (i.e. native texts of the same genre collected in the NKRY), undertaken in order to identify lexical differences between the three text varieties, was the focus of corpus research by Kunilovskaya et al.⁶⁰ The study was grounded in the concept of translation universals. Using selected quantitative indicators of style (sentence length, frequencies of morphological forms of word classes, lexical variety measured by the type/token ratio and the proportion of high-frequency words, and lexical density measured by the ratio of content words to the total number of words in texts) and suitable tests of statistical significance and effect size metrics⁶¹ found that, first, novice translations, extracted from the aforementioned Russian Learner Translator Corpus,⁶² differ more from native texts than translations performed by professionals (published by 10 selected Russian news portals), and, second, that both novice and professional translations revealed characteristics that made them linguistically different from native, non-translational texts (in terms of proportion of high frequency words, lexical density and specific frequency of

⁵⁹ M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian: Lexical profiles*, "Translation, and Interpreting" 2018, no. 10(1), pp. 33–52, <https://trans-int.org/index.php/transint/article/view/585/304> (30.08.2018); M. Kunilovskaya, and E. Lapshinova-Koltunski, *Translationese Features as Indicators of Quality in English-Russian Human Translation*, in: I. Temnikova, C. Orasan, G. Corpas Pastor., and R. Mitkov (Eds.), *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019)*, Varna, Bulgaria, September 5–6 2019, pp. 47–56, <https://www.aclweb.org/anthology/W19-8706.pdf> (30.08.20210).

⁶⁰ M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian...*; M. Kunilovskaya, and E. Lapshinova-Koltunski, *Translationese Features as Indicators of Quality...*

⁶¹ M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian: Lexical profiles...*

⁶² A. Kutuzov, and M. Kunilovskaya, *Russian Learner Translator Corpus...*

word forms).⁶³ In future research, it would be worthwhile verifying whether the same tendencies are observed in language pairs other than English-to-Russian, for example in the Polish-to-Russian translation conducted between more typologically similar Slavic languages. This and other avenues for future research will be discussed in the following Conclusions section.

The findings of this research are complemented by Kunilovskaya, and Lapshinova-Koltunski,⁶⁴ a study which compares originally produced vs. translated texts and good vs. bad students translations in terms of a theoretically motivated set of features. This set of features is based on the literature in the field of register studies that are known to capture translationese, supplemented by a number of morphosyntactic features that are meant to reflect fluency. The proposed features are a combination of relatively shallow lexicogrammatical indicators and of features that could be automatically tagged by a Universal Dependencies tagger, 45 in all. For instance, the authors included seven Universal Dependency features that have been shown to be translationese indicators specifically for the Russian-English pair: adjectival clause, auxiliary, passive voice auxiliary, clausal complement, subject of a passive transformation, asyndeton, a predicative or clausal complement without its own subject. In addition to the student translations from the RuLTC corpus and the professional translations from a range of established electronic media, the authors used original Russian newspaper texts as a reference corpus. The student translations were additionally evaluated for quality by translation experts, with only the top three best ranking and the bottom three worst rankings used in the analysis. For the study, the authors used a supervised learning model to solve two classification tasks based on the extracted feature frequencies: (1) translated vs. non-translated texts, and (2) best-ranking vs. worst-ranking translations. Overall, Kunilovskaya, and Lapshinova-Koltunski⁶⁵ found that the chosen features can distinguish between translated and non-translated texts with near perfect accuracy, that is, that these features reflect translationese reliably. The quality task, on the other hand, was not solved suc-

⁶³ M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian...*

⁶⁴ M. Kunilovskaya, and E. Lapshinova-Koltunski, *Translationese Features as Indicators of Quality...*

⁶⁵ *Ibidem*.

cessfully: the algorithm performed barely above the chance level. The authors conclude that translationese does not equal a reduction in quality; this might be due to the fact that ‘quality’ is more than lexico-grammatical faithfulness to standard; it reflects semantic faithfulness and pragmatic acceptability as well.⁶⁶

While the two studies we have just described in detail concern the newspaper register, other research demonstrates feasibility of corpus investigations into translationese in Russian in other genres. For example, Krasnopeeva⁶⁷, cited by Kunilovskaya et al.⁶⁸, looked at originally authored and translated Russian fiction and found statistically significant differences in lexis between the two. It is a promising avenue of research to (a) continue looking at the lexical, morphological and syntactic features distinguishing translated and non-translated Russian on the basis of large corpora; (b) continue looking into morphological, syntactic and textual features that impact register variation in translation;⁶⁹ and (c) to conduct corpus-based variationist research into “constrained language”⁷⁰ taking into account variables other than translation status. Such variables include, for example, translator’s proficiency (investigated in the two studies above), but also the setting for the translation/interpreting process, the communication medium, or the translation’s target audience.

5. CONCLUSIONS

In this paper, we focused on recent advances in corpus linguistic research on translation and interpreting, with particular emphasis on studies conducted using Russian language corpora. It goes without saying that there are other theoretical paradigms beyond empirical

⁶⁶ Ibidem.

⁶⁷ E. Krasnopeeva, *Lexical features of Russian translated discourse. A corpus-based comparative study of contemporary narrative prose*. Unpublished PhD dissertation, Chelyabinsk State University 2015 (cited in Kunilovskaya et al. 2018, pp. 34).

⁶⁸ M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian...*

⁶⁹ M. Kunilovskaya, and G. Corpas Pastor, *Translationese and Register Variation in English-To-Russian Professional Translation*, in: Vincent Wang, Lily Lim, and Defeng Li (Eds.). *New Perspectives on Corpus Translation Studies. New Frontiers in Translation Studies*, Springer, Singapore 2021 (https://doi.org/10.1007/978-981-16-4918-9_6).

⁷⁰ I. Lanstyak, and P. Heltai, *Universals in Language Contact and Translation*, “Across Languages and Cultures” 2012, 13(1), pp. 99–121.

corpus linguistic approaches that contribute to a better understanding of translation and interpreting. Our critical overview revealed first and foremost a dearth of freely available corpus resources in interpreting that include a Russian component - the problem that can be traced back to the practical difficulty of creating interpreting corpora. In addition, the EUROPARL compendium, which serves as a source for most existing interpreting corpora, is not helpful in the case of Russian.

Although Russian corpus linguistics is a fruitful discipline that has existed since the 1980s, it is possible to identify new directions and research gaps in the studies of translation and interpreting that may be pursued further in the near future. One of the areas that has enjoyed considerable popularity in recent years is research on so-called constrained communication, where language use is mediated by translation/interpreting, foreign language use or both,⁷¹ which is an underexplored area of study in the context of Russian language. Also, given the growing popularity of research on amateur, post-edited and L2 translation⁷² (e.g. into Russian as a foreign language for translators/interpreters) conducted on main European languages, we may note — at least to the knowledge of the Authors — the scarcity of such corpora of Russian, notably in the Polish-Russian language pair. Likewise, there are no Russian translation learner corpora with a Polish component, or Russian translation corpora comprising professional and amateur translations with a Polish component. As in much of academic research, translation/interpreting corpora including English are overrepresented among available resources: for the English-Russian language pair, there exist e.g. Kutuzov, and Kunilovskaya; Dayter;⁷³ Kunilovskaya, and Lapshinova-Koltunski.⁷⁴ Thus, research into translation taking into account the translator qualifications represents an important research gap for languages other than English. This gap could be addressed in the future provided that suit-

⁷¹ E.g. I. Lanstyak, and P. Heltai, *Universals in Language Contact...*; H. Kruger & B. Van Rooy, *Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English*, “English World-Wide” 2016, no. 37(1), pp. 26–57; M. Kajzer-Wietrzny, and Ł. Grabowski, *Formulaicity in constrained communication...*

⁷² This point has been also emphasized by Lefer (2021).

⁷³ A. Kutuzov, and M. Kunilovskaya, *Russian Learner Translator Corpus...*; D. Dayter, *Describing lexical patterns...*

⁷⁴ M. Kunilovskaya, and E. Lapshinova-Koltunski, *Translationese Features as Indicators of Quality...*

able corpora, representing those underexplored translation varieties, are developed. This may also contribute to the advancement of applied research on translation teaching and translation quality assessment, among others, in the Polish-Russian language pair.

Finally, it may be expected that more research will be conducted in the future using a variety of supervised or unsupervised machine learning techniques to study translational and interpreting phenomena through identification of hidden patterns of language use in large data collections. For example, Ustaszewski⁷⁵ showed how to apply supervised machine learning techniques (more precisely, Support Vector Machines (SVM) used for data classification) in order to computationally distinguish between direct translations from indirect translations based on a set of 26 linguistic features (i.e. represented as multidimensional vectors of numerical features). It is also clear that the growing popularity of machine learning techniques in the translation industry (deep learning, neural machine translation etc.) will have implications not only for the methodologies used by researchers studying translational phenomena, but also for the future of translation education.⁷⁶ This and other considerations promise a rise in popularity of interdisciplinary research and synergies between linguistics, literary studies, information technology / natural language processing, psychology etc. Advanced statistical and machine learning techniques will be brought to bear on corpus data to cast more light on the specificity of translation and interpreting, from the perspective of the product (text), process and function. These are only beginning to be explored for Russian, for example, in the work of Kutuzov, and Kuzmenko,⁷⁷ who used neural embedding models to study diachronic change. Collaborations between corpus and computational linguists may lead to new exciting research in the classic areas of interest we had identified in this paper.

In the same vein, we expect that there will be many more multifactorial studies⁷⁸ conducted in the future in Polish-Russian or Russian-

⁷⁵ M. Ustaszewski (2021) also presents important methodological considerations.

⁷⁶ G. Massey, and M. Ehrensberger-Dow, *Machine learning: Implications for translator education*. "Lebende Sprachen" 2017, no. 62(2), pp. 300–312.

⁷⁷ A. Kutuzov, and E. Kuzmenko, *Two centuries in two thousand words. Neural embedding models in detecting diachronic lexical changes*, in: Mikhail Kopotev, Olga Lyashevskaya, and Alto Mustajoki (Eds.), *Quantitative Approaches to the Russian Language*, Routledge, London 2017.

⁷⁸ See Paquot, and Gries (2020), notably chapters collected in Part V, for a more detailed discussion of statistical techniques.

Polish translation/interpreting research. Multifactorial approaches would enable one to posit explanatory hypotheses as to the text-external factors that govern the use of particular linguistic features (including, for example, particular translation equivalents) of translated, non-translated texts (native) or constrained texts. Monofactorial approaches that have been mostly applied so far in the studies of translationese are largely limited to the analysis (i.e. identification and description) of frequency distributions of particular linguistic features and of a single independent explanatory variable, and which, as such, only enable descriptive hypotheses. This future research direction, however, depends on whether researchers first develop suitable Polish-Russian translational and/or interpreting corpora, rich in metadata⁷⁹ (e.g. direction of translation, mode of delivery, text variety etc.), which is a *sine qua non* condition for such studies to flourish.

REFERENCES

- Baker, Mona. "Corpus Linguistic and Translation Studies: Implications and applications." *Text and Technology: in Honour of John Sinclair*. Baker, Mona, and Francis, Gill, and Tognini-Bonelli, Elena (Eds.). Amsterdam: John Benjamins, 1993: 233–250.
- Baker, Mona. "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research." *Target* 1995, no. 7(2): 223–243.
- Balakin, Alexey, and Bodrova, Alina, and Orekhov, Sergey, and Rybina, Maria. Russian-French Poetry Corpus 1800–1820 < <http://nevmenandr.net/fr/>>.
- Bernardini, Silvia, and Ferraresi, Adriano, and Miličević, Maja. "From EPIC to EP-TIC – Exploring simplification in interpreting and translation from an intermodal perspective." *Target* 2016, no. 28: 61–86.
- Calzada Pérez, Maria, and Laviosa, Sara. "Twenty-five years on: Time to pause for a new agenda for CTIS." *Reflexión crítica en los estudios de traducción basados en corpus / CTS spring-cleaning: A critical reflection*. Calzada Pérez, Maria, and Laviosa, Sara (Eds.). *MonTI* 13, 2021: 7–32.
- Čermák, František, and Rosen, Alexandr. "The case of InterCorp, a multilingual parallel corpus." *International Journal of Corpus Linguistics*, 2012, no. 13(3): 411–427.
- Chernov, Ghelly. *Inference and Anticipation in Simultaneous Interpreting*. Amsterdam: Benjamins, 1978/2004.
- Chesterman, Andrew. "Hypothesis about translation universals." Hansen, Gyde, and Malmkjær, Kirsten, and Gile, Daniel (Eds.). *Claims, Changes and Challenges in Translation Studies*. Amsterdam: John Benjamins, 2004: 1–13.

⁷⁹ The European Parliament Translation and Interpreting Corpus (EPTIC) (Bernardini et al. 2016) or TransBank corpus (Stauder, and Ustaszewski 2020) are examples of translational/interpreting corpora rich in metadata.

- Dayter, Daria. "Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN)." *FORUM* 2018, no. 16(2): 241–264 <<https://doi.org/10.1075/forum.17004.day>>.
- Dayter, Daria. "Collocations in non-interpreted and simultaneously interpreted English: a corpus study." *New Empirical Perspectives on Translation and Interpreting*. Vandevorde, Lore, and Daems, Joke, and Defrancq, Bart (Eds). London: Routledge, 2020: 67–91.
- Dayter, Daria. "Strategies in a corpus of simultaneous interpreting. Effects of directionality, phraseological complexity, and position in speech event." *Meta: Translator's Journal* 2021, no. 66(3): 594–617.
- Delaere, Isabel, and De Sutter, Gert. "Applying a Multidimensional, Register-Sensitive Approach to Visualize Normalization in Translated and Non-Translated Dutch." *Belgian Journal of Linguistics* 2013, no. 27: 43–60.
- De Sutter, Gert, and Lefer, Marie-Aude. "On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach." *Perspectives* 2020, no 28(1): 1–23.
- Egbert, Jesse, and Larsson, Tove, and Biber, Douglas. *Doing linguistics with a corpus: Methodological considerations for the everyday user*. [Cambridge Elements in Corpus Linguistics]. Cambridge: Cambridge University Press, 2020.
- Gorshkova, Anna. "Sopostavitel'noye issledovaniye sposobov dostizheniya ekvivalentnosti v sinkhronnom i pis'mennom perevodakh." Unpublished Phd dissertation, Moskva: Rossiyskiy universitet druzhby narodov, 2004. [Горшкова, Анна. "Сопоставительное исследование способов достижения эквивалентности в синхронном и письменном переводах." Unpublished PhD dissertation, Москва: Российский университет дружбы народов, 2004].
- Grabowski, Łukasz. "Corpora and Descriptive Translation Studies." *Ways to Translation*. Bogucki, Łukasz, and Goźdz-Roszkowski, Stanisław, and Stalmaszczyk, Piotr (Eds.). Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2015: 161–188.
- Gries, Stefan. "The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models." *Corpora* 2015, no. 10(1): 95–125.
- Gries, Stefan. "On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally." *Journal of Second Language Studies*, 2018, no. 1(2): 276–308.
- Gurin, Il'ya. Priemy rechevoy kompressii prisinkhronnom perevozhde s russkogo yazyka na angliyskiy. Unpublished Phd dissertation, Moskva: Moskovskiy gosudarstvennyy universitet, 2009. [Гурин, Илья. Приемы речевой компрессии при синхронном переводе с русского языка на английский. Unpublished PhD dissertation, Москва: Московский государственный университет, 2009].
- Herrmann, Berenike, and van Dalen-Oskam, Karina, and Schöch, Christoph. "Revisiting Style, a Key Concept in Literary Studies." *Journal of Literary Theory* 2015, no. 9(1): 25–52.
- Holmes, James. "The Name and Nature of Translation Studies." *The Translation Studies Reader*. Venuti, Lawrence (Ed.). London: Routledge, 1988/2000: 172–185.
- Ilukhin, Vladimir. *Strategii v sinkhronnom perevode*. Unpublished PhD dissertation, Moscow: 2001.
- Kajzer-Wietrzny, Marta, and Grabowski, Łukasz. "Formulaicity in constrained communication: An intermodal approach." *MonTI* 13, 2021: 148–183.

- Kalina, Sylvia. *Strategische Prozesse beim Dolmetschen: theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*. Tübingen: Günter Narr, 1998.
- Kisiel, Anna, and Koseska-Toszewa, Violetta, and Kotsyba, Natalia, and Satola-Staskowiak, Joanna, and Sosnowski, Wojciech. "Polish-Bulgarian-Russian Parallel Corpus". CLARIN-PL digital repository, 2016 <<http://hdl.handle.net/11321/308>>.
- Kluyeva, Natalia, and Bojar, Ondřej. "UMC o.1: Czech-Russian-English Multilingual Corpus". *Proceedings of International Conference Corpus Linguistics*, October 2008: 188–195 <<https://ufal.mff.cuni.cz/umc/cer/download.php?f=umc-o.1-paper-2008.pdf>>.
- Kopotev, Mikhail, and Mustajoki Arto, and Bonch-Osmolovskaya Anastasia (2021) "Corpora in Text-Based Russian Studies". *The Palgrave Handbook of Digital Russia Studies*. Daria Gritsenko, Mariëlle Wijermars, and Mikhail Kopotev (Eds). Palgrave Macmillan, Cham: 299–317 <https://doi.org/10.1007/978-3-030428556_17>.
- Kotze, Haidee. "Converging what and how to find out why: An outlook on empirical translation studies." *New Empirical Perspectives on Translation and Interpretation*. Vandevoorde, Lore, and Daems, Joke, and Defrancq, Bart (Eds.). Abingdon: Routledge, 2020: 333–370.
- Krasnopeeveva, Ekaterina. *Lexical features of Russian translated discourse. A corpus-based comparative study of contemporary narrative prose*. Unpublished PhD dissertation: Chelyabinsk State University, 2015.
- Kruger, Haidee, and Van Rooy, Bertus "Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English." *English World-Wide* 2016, 37(1): 26–57.
- Kunilovskaya, Maria, and Morgoun, Natalia, and Pariy, Alexey. "Learner vs. professional translations into Russian: Lexical profiles". *Translation, and Interpretation*, 2018, no. 10(1): 33–52 <<https://trans-int.org/index.php/transint/article/view/585/304>>.
- Kunilovskaya, Maria, and Lapshinova-Koltunski, Ekaterina. "Translationese Features as Indicators of Quality in English-Russian Human Translation." *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019)*. Temnikova, Irina, and Orasan, Constantin, and Corpas Pastor, Gloria, and Mitkov, Ruslan (Eds.). Varna, Bulgaria, September 5–6 2019: 47–56 <<https://www.aclweb.org/anthology/W19-8706.pdf>>.
- Kunilovskaya, Maria, and Corpas Pastor, Gloria. "Translationese and Register Variation in English-To-Russian Professional Translation." *New Perspectives on Corpus Translation Studies. New Frontiers in Translation Studies* Wang, Vincent, and Lim, Lily, and Li, Defeng (Eds.). Springer, Singapore, 2021 <https://doi.org/10.1007/978-981-16-4918-9_6>.
- Kutuzov, Andrey, and Kunilovskaya Maria. "Russian Learner Translator Corpus." *Text, Speech and Dialogue*. TSD 2014. Lecture Notes in Computer Science. Sojka Petr, and Horák, Aleš, and Kopeček, Ivan, and Pala, Karel (Eds.), vol. 8655. Springer: Cham, 2014 <https://doi.org/10.1007/978-3-319-10816-2_39>.
- Kutuzov, Andrey, and Kuzmenko, Elizaveta. "Two centuries in two thousand words. Neural embedding models in detecting diachronic lexical changes." *Quantitative Approaches to the Russian Language*. Kopotev, Mikhail, and Lyashevskaya, Olga, and Mustajoki, Alto (Eds.). London: Routledge, 2017.

- Lanstyak, Istvan, and Heltai, Pal. "Universals in Language Contact and Translation." *Across Languages and Cultures* 2012, no. 13:1: 99–121.
- Larsson, Tove, and Egbert, Jesse, and Biber, Douglas. "On the status of statistical reporting versus linguistic description in corpus linguistics: A ten-year perspective". *Corpora* 2022, no. 17(1), forthcoming.
- Laviosa, Sara. *Corpus-based translation studies: theory, findings, applications*. Amsterdam: Rodopi, 2002.
- Lederer, Marianne. *La traduction simultanée. Expérience et théorie*. Paris: Lettres modernes Minard, 1981.
- Lefer, Marie-Aude. "Parallel Corpora." *A Practical Handbook of Corpus Linguistics*. Paquot, Magali, and Gries, Stefan (Eds.). Cham: Springer Nature Switzerland AG, 2020: 257–282.
- Lefer, Marie-Aude. "Breaking new ground in contrastive and translation studies: Learner translation corpora to the fore". Plenary lecture delivered at the UCCTS 2021 conference in Bertinoro, Italy. 9–11 September 2021.
- Łaziński, Marek, and Kuratczyk, Magdalena, and Orekhov, Boris, and Słobodjan, Elena. "The Polish-Russian Parallel Corpus and Its Application in the Linguistic Analysis." *Prace Filologiczne*, 2012, vol. LXIII: 209–218.
- Łaziński, Marek, and Kuratczyk, Magdalena. "Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego." *Polskojęzyczne korpusy równoległe*. Gruszczyńska, Ewa, and Leńko-Szymańska, Agnieszka (Eds), Warszawa: Wydawnictwo Uniwersytetu Warszawskiego, 2016: 83–96.
- Łaziński, Marek, and Kuratczyk, Magdalena. "Pol'sko-russkiy parallel'nyy korpus Varshavskogo universiteta." *Język i Metoda* 2016, no. 3: 85–93 [Łaziński, Marek, and Kuratczyk, Magdalena. "Польско-русский параллельный корпус Варшавского университета." *Język i Metoda*, 2016, no 3: 85–93.]
- Malmkjær, Kirsten. "What Happened to God and the Angels: An Exercise in Translational Stylistics". *Target*, 2003, no. 15(1): 37–58.
- Massey, Gary, and Ehrensberger-Dow, Maureen. "Machine learning: Implications for translator education." *Lebende Sprachen* 62(2), 2017: 300–312.
- Mikhailov, Mikhail, Tommola, Hannu, and Isolhti, Nina. "Spoken to Spoken vs. Spoken to Written: Corpus Approach to Exploring Interpreting and Subtitling." *Polibits*, 2010, no. 41: 5–9.
- Mikhailov, Mikhail, and Cooper, Robert. *Corpus Linguistics for Translation and Contrastive Studies*. London: Routledge, 2016.
- Mikhailov, Mikhail, and Santalahti, Miia, and Souma, Julia (Eds). *PEST. A parallel electronic corpus of state treaties*. Amsterdam: John Benjamins, 2019.
- Neubert, Albrecht, and Shreve, Gregory. *Translation as Text*. Kent-London: The Kent State University Press, 1992.
- NKRYA. Natsional'nyy korpus russkogoyazyka 2003–2005. Moskva: Indrik, 2005. <<https://ruscorpora.ru/>> [НКРЯ. Национальный корпус русского языка 2003–2005. Москва: Indrik, 2005 <<https://ruscorpora.ru/>>].
- Olohan, Meave. *Introducing Corpora in Translation Studies*. Routledge: London, 2004.
- Orekhov, Boris "Parallel'nyy korpus perevodov «Slova o polku Igoreve»: itogi i perspektivy." *Natsional'nyy korpus russkogo yazyka: 2006–2008. Novyye rezul'taty i perspektivy*. Plungyan, Vladimir (Ed.). Sankt-Peterburg: Nestor-Istoriya, 2009: 462–473 [Орехов, Борис "Параллельный корпус переводов «Слова о полку Игореve»: итоги и перспективы." *Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы*.

- Плунгян, Владимир (Ed.). Санкт-Петербург: Нестор-История, 2009: 462–473].
- Paquot, Magali, and Gries, Stefan (Eds.). *A Practical Handbook of Corpus Linguistics*. Cham: Springer Nature Switzerland AG, 2020.
- Peżik, Piotr, and Ogrodniczuk, Maciej, and Przepiórkowski, Adam. *Parallel and spoken corpora in an open repository of Polish language resources. Human Language Technologies as a Challenge for Computer Science and Linguistics*. LTC Poznań 2011 <<http://pelcra.pl/new/polrus>>.
- Piao, Zhehao, and Li, Qinghua, and Wang, Lixia. “Jiyu han-e pingxing yuliaoku de “jiu guo” dieyinci eyi guili yanjiu.” *Zhongguo eyu jiaoxue*, 2014, no 3: 46–51.
- Polskoy, Sergey, and Rjéoutski, Vladislav. *The Corpus of Russian Translations*, 2021 <<https://krp.dhi-moskau.org/ru/translations>>.
- Rybicki, Jan. “Stylometric translator attribution: do translators leave lexical traces?” *The Translator and the Computer*. Piotrowski, T., and Grabowski, Ł. (Eds). Wrocław: Wydawnictwo Wyższej Szkoły Filologicznej, 2013: 193–204.
- Saldanha, Gabriela. “Translator style. Methodological considerations.” *The Translator*, 2011, no. 17(1): 25–50.
- Setton, Robin, and Adelina Hild. “Editors’ critical foreword.” *Inference and Anticipation in Simultaneous Interpreting*. Chernov, Ghelly (Ed.). Amsterdam: Benjamins, 2004: IX–XXII.
- Shiryayev, Anatolij. *Sinkhronnyj Perevod*. Moscow: Voenizdat, 1979.
- Shlesinger, Miriam. “Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies.” *Meta* 1998, no. 43(4): 486–493.
- Sitchinava, Dmitry. “Parallel Corpora within the Russian National Corpus.” *Prace Filologiczne* 2012, vol. LXIII: 271–278 <<https://courses.helsinki.fi/sites/default/files/course-material/4443508/CEEOL-4.pdf>>.
- Sönnig, Lukas, and Werner, Valentin. “The replication crisis, scientific revolutions, and linguistics.” *Linguistics*, 2019, no, 59 (5): 1179–1206.
- Stauder, Andy, and Ustaszewski, Michal. “TransBank: A Meta-Corpus for Translation Research.” *Translation in the Digital Age. Translation 4.0*. Sinner, Carsten, and Paasch-Kaiser, Christine, and Härtel, Johannes (Eds.), Newcastle upon Tyne: Cambridge Scholars Publishing, 2020: 69–80.
- Szabo, Martina, and Vincze, Veronika, and Nagy, Istvan. “HunOr: A Hungarian–Russian Parallel Corpus.” *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul: European Language Resources Association (ELRA), 2012: 2453–2458. <http://www.lrec-conf.org/proceedings/lrec2012/pdf/262_Paper.pdf>.
- Tiedemann, Joerg. “OPUS - Parallel Corpora for Everyone.” *Baltic Journal of Modern Computing* 2016, no. 4 (2). Special Issue: *Proceedings of the 19th Annual Conference of the European Association of Machine Translation (EAMT)*.
- Tymoczko, Maria. “Computerized Corpora and the Future of Translation Studies.” *Translators’ Journal*, 1998, no. 43(4): 652–660.
- Ustaszewski, Michal., and Stauder, Andy. “Corpora in Translation Studies: Why Do We Need Them and What Can Current Resources (Not) Offer to Us?.” *Translation in the Digital Age. Translation 4.0*. Sinner, Carsten, and Paasch-Kaiser, Christine, and Härtel, Johannes (Eds). Newcastle upon Tyne: Cambridge Scholars Publishing, 2020: 56–68.
- Ustaszewski, Michal. “Towards a machine learning approach to the analysis of indirect translation.” *Translation Studies*, ahead of print, 2021: 1–19. <<https://doi.org/10.1080/14781700.2021.1894226>>.

CORPUS-BASED AND CORPUS-DRIVEN RESEARCH...

- Vinay, Jean-Paul, and Darbelnet, Jean. *Comparative Stylistics of French and English: A methodology for translation* (translated and edited by Juan Sager, and Marie-Josée Hamel). Amsterdam: John Benjamins, 1958/1995.
- von Waldenfels, Ruprecht. "Compiling a parallel corpus of slavic languages. Text strategies, tools and the question of lemmatization in alignment." *Beiträge der Europäischen Slavistischen Linguistik: POLYSLAV 9*. Brehmer, Bernhard, and Zdanova, Vladislava, and Zimny, Rafal (Eds.). München: Verlag Otto Sagner, 2006: 123–138.
- Winter, Bodo. *Statistics for Linguists: An Introduction Using R*. London: Routledge, 2019.
- Zanettin, Federico. *Translation-Driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome Publishing, 2012.
- Ziemski, Michał, and Junczys-Downmunt, Marcin, and Pouliquen, Bruno. "The United Nations Parallel Corpus v1.0." *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16)*. Slovenia: Portorož, 2016: 3530–3534.
- Zimnyaya Irina, and Chernov Geliy. "Veroyatnostnoye prognozirovanye v protsesse sinkhronnogo perevoda." *Predvaritel'nyye materialy eksperimental'nykh issledovaniy po psikholingvistike*. Leont'yev, A.A., and Zhinkin, N.I., and Shakhnarovich, A.M. (Eds.). Moskva: Iya-ANSSSR, 1973: 110–116 [Зимняя Ирина, and Чернов Гелий. "Вероятностное прогнозирование в процессе синхронного перевода." Леонтьев, А.А., and Жинкин, Н.И., and Шахнарович, А.М. (Eds.). *Предварительные материалы экспериментальных исследований по психоллингвистике*. Москва: ИЯ-АН СССР, 1973: 110–116].

**ALEKSANDER KIKLEWICZ** ORCID: <https://orcid.org/0000-0002-6140-6368>

Uniwersytet Warmińsko-Mazurski w Olsztynie

КОРПУСНЫЕ ИССЛЕДОВАНИЯ В СИНТАКСИСЕ: ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ¹

CORPUS RESEARCH IN SYNTAX: OPPORTUNITIES AND LIMITATIONS

The aim of the article is to show the possibilities of using internet corpora in syntactic research. Corpus analysis allows for the verification of axiomatically modeled compositions of units, however the command may be material (exact search) or algorithmic (lexical-grammatical search). Corpus tools are primarily useful in determining the frequency of units and their combinations in texts, and the regularity of predetermined features. Based on the frequency, the typological features of different languages can be described in terms of selected parameters. Corpus analysis also makes it possible to check whether a given lexeme valence property is certified or not, and to determine its functionality. The author also shows that the use of the internet corpus in syntactic research has certain limitations. In the case of frequent phenomena, especially in colloquial speech, corpus analysis is effective, but it does not always allow to document phenomena that are less typical and stylistically marked. It is all the more dangerous in the case of less frequent, but regular units in terms of the implementation of systemic oppositions. The corpus should be equally capable of describing the phenomena of speech and the language system, concludes the author. One conclusion is that corpus analysis should be set up with introspection and polling.

ВВЕДЕНИЕ

В последние десятилетия корпусная лингвистика стала одним из наиболее популярных направлений в области языкознания, а обращение к корпусам стало обычной, рядовой практикой лингвистов. Существование больших и постоянно обновляемых интернет-картотек позволяет ускорить исследовательские процедуры,

¹ Автор благодарит своих коллег: проф. Марка Лазинского (Marek Łaziński) из Варшавского университета и Адама Пшепюрковского (Adam Przepiórkowski) из Института переработки информации ПАН, за полезные советы, касающиеся доступа к ресурсам Национального корпуса польского языка.

а также получить более качественную информацию о языковой системе и ее реализации в текстах. В то же время придется осознать некоторые ограничения корпусного метода. Корпус представляет собой определенную выборку и не в состоянии охватить все продукты языковой деятельности. Даже если бы мы вообразили максимально объемный корпус, включающий в себя все тексты, написанные на данном языке, в процессе создания корпуса будут появляться новые тексты, а часть элементов, отмеченных в корпусе, станет функционально нерелевантной.

Лингвистика в широком смысле имеет дело с описанием четырех категорий объектов, постулируемых в общей теории систем. В соответствии с концепцией Юрия Урманцева (1978: 10 и др.) существуют четыре параметра, необходимые для описания каждой системы: 1) первичные элементы; 2) их свойства как основания для категоризации; 3) отношения единства, благодаря которым элементы образуют классы; 4) законы композиции, благодаря которым из элементов строятся конструкции. Эмпирическое исследование призвано предоставить исходную информацию о некоторых из этих параметров, а именно — осуществить инвентаризацию языковых элементов и всех отмеченных композиций. Языковой корпус в интернете идеально подходит для решения этих задач, неслучайно он получил широкое употребление в исследованиях синтаксиса. В данной статье обсуждаются некоторые актуальные проблемы синтаксического анализа с использованием корпусных данных. Главное внимание будет посвящено эмпирическим аспектам реализации модели экспликативного синтаксиса, разрабатываемой группой исследователей, относящихся к польской школе семантического синтаксиса. Основная часть иллюстративного материала получена из Национального корпуса русского языка (<https://ruscorpora.ru/>) — ссылки на корпус имеют вид аббревиатуры НКРЯ.

1. КОРПУСНЫЕ ИССЛЕДОВАНИЯ В ОБЛАСТИ СИНТАКСИСА: КРАТКИЙ ОБЗОР ПРОБЛЕМАТИКИ

Роль интернет-корпусов особенно важна в случае аксиоматических описаний, когда на основе ранее принятых постулатов генерируются абстрактные модели, например, типовые конструк-

ции единиц, реальность, а также функциональность которых следует установить посредством наблюдений. Корпус позволяет введение параметров, по которым компьютерная система осуществляет поиск необходимых единиц (токенов). Эти параметры могут быть введены разными способами. Первый из них имеет материальный характер и опирается на языковую интуицию. Речь идет о ситуациях, когда исследователь, благодаря своей языковой компетенции, генерирует некоторые возможные реализации модели — обращение к корпусу в этом случае должно подтвердить существование гипотетических конструкций². К примеру, задача исследователя может касаться обоснования синтаксического статуса глагола *содействовать* как двухместного предиката второго порядка. В четвертом томе *Словаря русского языка* (Евгеньева 1984: 180) в соответствующей статье приводятся два примера:

- (1) Обстоятельство это [...] содействовало его успеху.
- (2) Наступлению наземных войск активно содействовали удары нашей авиации.

Как видим, синтаксические коллокации глагола — абстрактные существительные *обстоятельство, успех, наступление, удары*. Статус предиката второго порядка (за которым стоит выражаемое данным словом понятие) требует, чтобы его пропозициональные аргументы (указывающие на ситуации, события, состояния, процессы и отношения) были выражены наиболее полным, дискретным, изосемическим способом, т.е. в форме придаточного предложения. Если мы постулируем, что в содержании глагола *содействовать* закодирована пропозициональная структура $P(p, q)$, необходимо привести отмеченные формы синтаксических актантов в первой и второй позиции. Поиск-вый параметр может быть определен с учетом знания о том, как оформляются в русском языке придаточные изъяснительные предложения. Рассмотрим это на примере использования опции «Поиск точных форм» в НКРЯ:

- (3) поисковый параметр: [содействовал тому, что]
Александр содействовал тому, что войска обесславили себя вторжением в Испанию (НКРЯ).

² Михаил Котин (Kotin 2018: 63) пишет, что корпусный анализ должен сопровождаться самоанализом как критерием языковой компетенции.

- (4) поисковый параметр: [содействовало то, что]
Этому содействовало то, что Платов [...] был тесно связан с донским казачеством (НКРЯ).

Материальный параметр в случае синтаксического анализа имеет свои ограничения: предложение, кроме известного исследователю глагольного предиката и синтаксических операторов, содержит определенный лексический, трудно прогнозируемый состав. В этом случае точный поиск невозможен и следует прибегнуть к другой форме — лексико-грамматическому поиску, используя проведенную в текстах разметку. Эта опция необходима для поиска предложений, в которых оба пропозициональных аргумента при предикате *содействовать* выражены в форме придаточных предложений. Запрос имеет вид:

[что] [содействовать V,indic] [что]

Компьютерная программа открывает единственный имеющийся в корпусе пример — из прозы Льва Толстого:

- (5) Тому, что Багратион выбран был героем в Москве, содействовало и то, что он не имел связей в Москве, и был чужой. (НКРЯ)

Интерпретация этого факта — это уже дело исследователя. В зависимости от своих теоретических предпосылок он считает возможным трактовать этот факт как подтверждение сочетаемости глагола с сентенциальным актантом или продолжит поиск в других источниках. Например, если в поисковой системе Google вписать поисковый параметр "то что * содействовало тому что", можно получить доступ к единственному предложению:

- (6) То, что Шелеста сняли с постов, содействовало тому, что на рубеже 1980-х — 1990-х население Центральной Украины не имело уже таких эффективных лидеров и руководителей как Шелест³.

Несмотря на то, что конструкция с сентенциальными актантами представлена единичными примерами использования, они вполне соответствуют как литературной норме, так и системе современного русского языка.

³ <https://pandia.ru/text/77/463/20680-3.php> (16.12.2021).

Как видим, интуиция или интроспекция является не только альтернативным источником лингвистической информации по отношению к текстам, в которых языковые единицы и их конструкции считаются отмеченными, — в определенной степени она также влияет на поисковые процедуры при работе с интернет-корпусом (подробнее об этом см.: Meurers 2005). Другими словами, использование корпуса предполагает наличие некоторых предварительных языковых знаний субъекта, что особенно важно в ситуациях, когда корпус располагает ограниченными возможностями автоматизированного поиска. Так, Арто Мустайоки (Mustajoki 2006: 53 и др.) рассматривает возможности корпусного исследования русских безличных предложений типа

Лодку унесло ветром.

В связи с тем, что используемый исследователем корпус Integrum⁴ не содержит грамматической разметки материала, приходится использовать дополнительные инструменты, чтобы сузить поле поиска. Мустайоки принимает во внимание факт, что в безличных предложениях в позиции каузатора выступают определенные глагольные лексемы: *унесло, убило, зарезало, ударило* и др. Указав эти лексемы в поисковом запросе, исследователь смог ограничить поиск необходимых конструкций. В тех же случаях, когда общее количество конструкций с некоторыми единицами было слишком большим, выбирались случайные примеры. Таким образом был получен подкорпус, состоящий из 2304 предложений, которые были экспортированы в программу Excel и подверглись дальнейшему лингво-статистическому анализу.

Разумеется, поиск в корпусе может иметь сугубо алгоритмический характер, как в случаях, когда нам известны свойства матричной конструкции, а обращение к корпусу должно показать их отмеченность или неотмеченность, а также (в случае отмеченности) частоту употребления. Так, в окне синтаксического поиска НКРЯ⁵ мы можем вписать:

[V,сов] сочин [V,несов]
[V,несов] сочин [V,сов].

⁴ <http://www.integrumworld.com>

⁵ <https://ruscorpora.ru/new/search-syntax.html> (16.08.2021)

Такая команда означает, что мы заинтересованы в нахождении сочинительных конструкций с двумя глаголами, один из которых выражает значение совершенного, а другой — значение несовершенного вида. Программа открывает нам языковые выражения, удовлетворяющие этому требованию, например:

- (8) Охрана перепугалась, говорит: подходить нельзя.
 Vpf Vipf
- (9) Он спасся, пытался плыть, ныряя под воду.
 Vpf Vipf
- (10) Зачем много зарабатывать, стать культурным человеком?
 Vipf Vpf
- (11) Краны на кухне даже не сипели, превратились в воспоминание.
 Vipf Vpf

По данным корпуса, первый тип является более регулярным: он представлен 35 единицами, тогда как второй — только 21 единицей, что дает основание для некоторых первичных обобщений. Важно то, что конструкции, в которых глаголы согласованы по виду, представлены в значительно большем объеме, при этом несовершенный вид употребляется чаще:

[V,несов] сочин [V,несов] — 536 предложений
 [V,сов] сочин [V,сов] — 358 предложений

По мнению специалистов (см. Biber, Fitzmaurice, Reppen 2002; Gries 2009; Mustajoki 2006; Münzberg 2011; Marzo, Heylen, Sutter 2012), корпусные исследования чрезвычайно эффективны при изучении функциональности тех или иных синтаксических явлений, в частности, при определении частоты их употребления. Лукаш Пёнтковский (Piątkowski 2018: 133) показывает, что такое исследование может быть проведено при сопоставлении языков, в частности, немецких и польских вербономинальных конструкций типа:

- (12) польск. brać udział ‘принять/принимать участие’
 (13) польск. przynosić/przynieść efekt ‘дать/давать эффект’
 (14) нем. einen Besuch machen ‘навестить/навещать’
 (15) нем. Maßnahmen treffen ‘принять/принимать меры’

Благодаря корпусным данным определяются левые и правые коллокаты каждой рассматриваемой конструкции, а их стати-

стическая обработка и обобщение позволяют получить информацию о типовой сочетаемости вербономинальных фразем.

Таблица 1. Типовая схема сочетаемости немецкого словосочетания
Maßnahmen treffen

подлежащее		обстоятельство	определение		определение/ дополнение
Regierung	trifft	bereits rechtzeitig unverzüglich	(die)erfolgreichen (die) notwendigen geeignete	Maßnahmen	zum Schutz zur Sicherheit zur Verbesserung

Таблица 2. Типовая схема сочетаемости польского словосочетания
podjąć kroki

подлежащее		определение	обстоятельство/ определение	
rząd firma	podjęła	kroki	prawne zaradcze	w celu w kierunku przeciwko

Применение корпусных данных при сопоставлении синтаксиса языков можно показать на примере русских и польских конструкций с адъективными (в частности указательными) местоимениями. Предварительные наблюдения свидетельствуют, что в польском языке чаще выступает обратный порядок слов, т.е. конструкции, в которых существительное находится в препозиции по отношению к определению. На это указывают некоторые примеры из параллельного корпуса⁶:

- (16) польск. Wywalców *fakt ten* nie zbulwersował jednak (Andrzej Sapkowski).
[перевод] Завседгатаев *этот факт*, однако, не взволновал и не мешал им.
- (17) *Problem ten*, przez metodologów uznany pospiesznie za metafizyczny, tlił się na dnie wszystkich niemal dyskusji i sporów (Stanisław Lem).
[перевод] *Эта проблема*, поспешно признанная метафизической, тлела на дне всех дискуссий и споров.

Разумеется, в каждом языке имеются конструкции с прямым и обратным порядком слов, однако в соответствии с нашей гипотезой в польском языке инверсия представляет собой более регулярное явление. Обратимся к НКРЯ и воспользуемся лек-

⁶ <https://processing.ruscorpora.ru/>

сико-грамматическим поиском. В данном случае следует исключить из рассмотрения употребление указательного местоимения *это* в форме среднего рода, ввиду его регулярного использования в позиции связки, т.е. в рамках другого (пролептического) отношения. Вот результат поиска:

[APRO,nom,(m|f)] [S,nom] – 320.497 (84%)
 [S,nom] [APRO,nom,(m|f)] – 61.921 (16%)

В Национальном корпусе польского языка (в версии, разработанной Институтом переработки информации ПАН) поиск осуществляется с помощью команд, которые выглядят несколько по-другому:

[pos=subst & case=nom] [orth=»ten|ta|te|ci»]
 [orth=»ten|ta|te|ci»] [pos=subst & case=nom]

В сбалансированном корпусе (объемом 300 млн сегментов) имеется 161.323 конструкции с прямым порядком слов (что составляет 65%) и 87.092 конструкции с обратным порядком слов (35%). Как видим, и в польском материале конструкции с прямым порядком доминируют, однако здесь инверсия охватывает 1/3 всех сочетаний, тогда как в русском языке употребительность этого явления в два раза меньше.

2. КОРПУСНЫЙ МЕТОД В МОДЕЛИ ЭКСПЛИКАТИВНОГО СИНТАКСИСА

Концепция экспликативного синтаксиса была разработана в публикациях Станислава Кароляка (Karolak 1984; 2002), а также дополнена в работах его последователей (Kiklewicz, Korytkowska 2010; Kiklewicz, Korytkowska, Mazurkiewicz-Sułkowska et al. 2019). Предметом исследований являются определяемые в терминах грамматических классов синтаксические структуры, рассматриваемые как репрезентации исходных семантических пропозиций, представляющих конфигурации вершинного (матричного) предиката и его компонентов, обусловленных заключенным в предикате лексическим понятием. Одной из форм представления синтаксических репрезентаций являются, например, экспликативные схемы следующего вида:

- (18) $P(x, q) \rightarrow V_p - N_x - S_q$
 Людмила Николаевна знала, что один лишь эгоизм движет
 поступками Виктора (НКРЯ).

Важнейшая особенность экспликативного синтаксиса, отличающая его от других моделей (например, московской семантической школы), состоит в том, что синтаксические формы упорядочиваются в соответствии с тем, насколько полно они отражают базовую пропозициональную структуру. Вслед за Галиной Золотовой это свойство определяется как изосемичность. В случае реализации пропозициональных аргументов, то есть таких, которые указывают на ситуации, события, состояния, процессы и отношения, в русском языке имеется целый ряд альтернативных форм выражения, более или менее соответствующих характеру пропозиционального аргумента. Например, глагол *знать* употребляется как двухместный предикат с пропозициональным аргументом во второй позиции. Этот аргумент может быть выражен несколькими способами. В НКРЯ возможные формы сочетаемости глагола с его правым коллокатом можно получить, вписав в окне лексико-грамматического поиска команду [знать V,indic]. Приведенный ниже список содержит формы выражения пропозиционального аргумента, в том числе полные, неполные (являющиеся следствием имплицитации) и сверхполные (являющиеся следствием расщепления):

$q \rightarrow S_q^7$	Эраст знает сам, что происходит в душе его.
$q \rightarrow S_q S_q$	О том, что произошло вчера, он знает, что это была провокация националистов.
$q \rightarrow NV_{aq} S_q$	О данном событии он знает, что его причины — в эскалации конфликта.
$q \rightarrow V_q NV_q$	О том, что произошло на площади, Федор знает все подробности.
$q \rightarrow NV_q$	Левон Оганезов знает множество историй.
$q \rightarrow NV_{aq} \emptyset_q$	Петя знает правильное решение задачи.
$q \rightarrow NV_{aq} NV_{aq} \emptyset_q$	О данном событии Иван знает все подробности.
$q \rightarrow NV_{aq} NV_{Pq} \emptyset_{Vq}$	Святитель Василий знал медицину как языческое искусство.
$q \rightarrow N_{aq} \emptyset_q$	Батюшка знает меня.
$q \rightarrow N_{aq} N_{Pq} \emptyset_{Vq}$	Любители поэзии знают Шульпякова как одного из последних переводчиков.
$q \rightarrow \emptyset_q$	[С Иваном не спорьте] — Иван знает.

Краткие синтаксические формы (например, именная группа в позиции пропозиционального аргумента) более распростра-

⁷ Определения символов см. в работе: Kiklewicz, Korytkowska, Mazurkiewicz-Sułkowska et al. 2019: 13 и др.

нены в речи, особенно в разговорном стиле, тогда как полные, изосемические формы имеют книжный, а иногда и достаточно искусственный характер. Именно в случае таких форм необходима верификация их отмеченности, т.е. подтверждение того, что данный тип экспликации аргумента предусмотрен валентностью глагола и реализуется в речи. Например, если мы обратимся к русскому глаголу *вздумать*, в первую очередь вспоминаются конструкции с инфинитивом. Одна из них — название народной сказки:

(19) Как волк *вздумал* хлеб печь.

N_x V_p V_{inf}

Языковая интуиция, однако, может быть недостаточно объективной при решении вопроса о том, возможна ли в этом случае сочетаемость глагола с сентенциальным актантом. Более надежным способом является обращение к корпусу, в котором можно найти единственный пример требуемой конструкции:

$q \rightarrow S_q$ А жадный купец *вздумал*, что в этой палке заключена чудесная сила.

N_x V_p S_q

Этот пример дает повод задуматься о том, насколько обращение к корпусу может помочь в решении проблемы. Приведенное выше предложение, с одной стороны, является лингвистическим фактом, но, с другой стороны, факт этот единичный и к тому же заимствован из художественного текста начала XX века. Чтобы убедиться в том, что такого рода сочетаемость характерна для современного русского языка, следует обратиться к другим источникам. Так, в поисковой системе Google можно найти следующие примеры:

(20) Ты же знаешь, как быстро вокруг нас рухнет всё общество, если Хан *вздумает*, что ему можно безнаказанно меня подкалывать?⁸

(21) Может, *вздумает*, что именно он замена Хави и останется?⁹

(22) И ежели кто *вздумает*, что планы его расходятся с Высшими Целями Гильдии Серых Магов, [...] он может считать себя в состоянии войны с Гильдией¹⁰.

⁸ <https://context.reverso.net/перевод/русский-иврит/вздумает> (16.12.2021).

⁹ <http://192.168.8.1/html/quicksetup.html> (16.12.2021).

¹⁰ <https://mikportal.org/forum/topic2058s75.html?#entry81724> (16.12.2021).

- (23) Разве не безумным сочтут слепца, который *вздумает*, что знает различия цветов, не зная цвета?¹¹
- (24) Пока не *вздумали*, что мы здесь короли!¹²
- (25) Дочь конюха и свиарки *вздумала*, что она принцесса!¹³

Несмотря на то, что НКРЯ содержит 40 млн словоупотреблений, и в нем представлены тексты разных стилей, в то же время, как констатирует Владимир Плунгян (2005), 40% текстов относятся к художественному стилю. Напротив, участие научных и официально-деловых текстов относительно невелико и, следует помнить, что полные и компрессированные синтаксические формы по-разному представлены в текстах разных стилей (Korytkowska, Małdziejewa 2002: 161; Kiklewicz 2016: 106). Михаил Котин (Kotin 2018: 63 и др.) справедливо отмечает, что, работая с корпусами, даже весьма обширными, исследователи часто озабочены адекватной фиксацией описываемых единиц и типов в их реальной текстовой среде. Также ученый пишет, что необходимо учитывать зависимость между частотой использования единицы и профилем текстов, образующих корпус, что непосредственно влияет на интерпретацию материала. Часто, по мнению Котина, возникают серьезные коллизии между результатами корпусного анализа и «текущим состоянием» в области категоризации языковых форм. Эти расхождения, по его мнению, дополнительно усиливаются из-за того, что недостаточно принимаются во внимание такие «экстракорпоральные» факторы, как, например, внешние условия речевой деятельности.

Указанные ограничения проявляются в процессе синтаксического анализа и, в частности, при верификации экспликативных схем. Рассмотрим в качестве примера глагол *интерпретировать*. С семантической точки зрения, он представляет собой трехместный предикат второго порядка: $P(x, p, q)$. В соответствии со значением глагола некоторое лицо интерпретирует некоторый факт, положение дел и т.д. как некоторый другой (более известный) факт или положение дел, т.е. ставит в соответствие одному пониманию положения дел другое его понимание. В НКРЯ можно найти большое число примеров того, что по-

¹¹ http://bookwu.net/book_filosofiya_677/23_tema-13 (16.12.2021).

¹² <https://stih.ru/2013/08/22/5244> (16.12.2021).

¹³ <https://litnet.com/ru/reader/kak-byvshaya-princessa-byt-nalazhivala-b359791?c=> (16.12.2021).

зиции *p* и *q* реализованы в форме абстрактных существительных (или соответствующих именных групп), например:

NS_q NS_r Он интерпретирует его замысел как попытку проникновения в историю вопроса.

В то же время мы не находим в корпусе ни одного примера того, что пропозициональный аргумент реализуется в форме придаточного предложения. На этом основании можно было бы констатировать, что данный глагол не предусматривает сочетаемость с сентенциальными актантами, хотя такой вывод был бы опрометчивым: характер полученных данных может быть слишком зависим от профиля собранных в корпусе текстов. И действительно, если мы обратимся к другим источникам, например, к поисковым системам интернета, окажется, что конструкции с сентенциальными актантами также существуют. Вот несколько примеров из поисковой системы Rambler.ru:

- S_q NS_r То, что происходит, можно интерпретировать как начало избирательной кампании¹⁴.
- NS_q S_r Совокупность предлагаемых норм можно интерпретировать как то, что для расчета прибыли КИК на основании финансовой отчетности КИК должна располагаться в стране-партнере РФ¹⁵.
- S_q S_r То, что нарисовано на картинке, можно интерпретировать как то, что «тирамису побеждает пончик с желе».
То, что они услышали, он интерпретирует как то, что Европа дает наказ правительству.
То, как ведет себя ЕС, можно интерпретировать как то, что Запад начал некое «тактическое отступление» по украинскому вопросу¹⁶.

О том, что приведенные выше синтаксические формы отражают реальные, закодированные в языковой системе валентные свойства глагола, свидетельствует не только их материальная отмеченность в текстах, но и их совместимость с лингвистической интуицией. В связи с этим можно сослаться на высказывание Казимежа Полянского, который придавал этому источнику лингвистической информации особое значение. В предисловии к дистрибутивно-генеративному словарю польских глаголов он писал:

¹⁴ <https://www.svoboda.org/a/28311562.html> (16.12.2021).

¹⁵ <https://docplayer.com/29940887-Podgotovlena-novaya-versiya-popravorov-v-deofshorizacionnoe-zakonodatelstvo.html> (16.12.2021).

¹⁶ Три последних примера из работы: Киклевич 2016: 16.

В соответствии с постулатом порождающей грамматики лингвистика не должна ограничиваться так называемыми текстовыми корпусами — в поисках лингвистических фактов очень часто приходится обращаться к собственной лингвистической интуиции и интуиции других пользователей языка¹⁷ (Polański 1980: 6).

Наиболее эффективным представляется комплексный метод регистрации языкового материала, т.е. с использованием текстовых источников, языковой интуиции, анкетирования, эксперимента, интернет-корпуса и поисковых систем интернета. Это можно показать на примере исследования конструкций русского языка с расщеплением пропозиционального аргумента, т.е. таких, в которых одной аргументной позиции (в семантической структуре) соответствуют два синтаксических актанта (в формально-грамматической структуре), например:

- (26) Он раздражает меня как факт, внезапно нарушивший наш покой (НКРЯ).
 || Меня раздражает то, что он представляет собой факт, внезапно нарушивший наш покой...
- (27) Меня раздражает белый карандаш как невероятная глупость (НКРЯ).
 || Меня раздражает то, что белый карандаш представляет собой невероятную глупость.

Глагол *раздражать* реализует пропозициональную схему $P(p, y)$, которая означает: ‘некоторое положение дел вызывает раздражение некоторого лица’. В первом предложении пропозициональный аргумент выражен элементами *он... как факт*, во втором предложении — элементами *белый карандаш... как невероятная глупость*.

Когда возникает вопрос о подобной сочетаемости другого глагола той же лексико-семантической группы — *удивлять*, мы находим в корпусе только несколько примеров из текстов XIX века:

- (28) Но и между произведениями древних, не большая ли часть удивляют нас как памятники высокого духа, а не как образцы совершенного вкуса.
- (29) Как люди, отжившие свой век, они удивляли и забавляли нас своей оригинальностью и разными причудами.

¹⁷ „Zgodnie z postulatem gramatyki generatywnej nie pozwalającym lingwiście ograniczać się do tzw. korpusu tekstowego trzeba w poszukiwaniu faktów językowych bardzo często sięgać do intuicji językowej własnej i innych współużytkowników języka”.

- (30) Как иностранец, он удивляет князя Владимира и его княгиню заморскими подарками.
 (31) Как общественный симпатом, — продолжал он, — это меня несколько не удивляет.

Возникает дилемма: дают ли эти несколько примеров текстов 150–200-летней давности основание утверждать, что *удивлять*, как и другие эмотивные глаголы, допускает расщепление пропозиционального аргумента? Лингвистическая интуиция подсказывает нам, что такие конструкции возможны и сегодня, однако, чтобы задокументировать их, нужно обратиться к другим источникам. В поисковой системе Rambler.ru можно найти интересные нас предложения с глаголом *удивлять*, причем они не являются единичными, ср.

- (32) Он меня удивляет, скорее, как чудо Природы.
 $N_p \quad N_y \quad V_p \quad N_p \quad \emptyset_p$
 (33) Меня это удивляет как факт и не удивляет как результат их работы.
 $N_y \quad N_p \quad V_p \quad N_p \quad \emptyset_p$
 (34) Появление [...] животных уже мало кого удивляет как явление.
 $NV_p \quad N_y \quad V_p \quad N_p \quad \emptyset_p$
 (35) Твоя энергетика [...] удивляет как феномен!
 $NS_p \quad V_p \quad \emptyset_y \quad N_p \quad \emptyset_p$

Таким образом, можно с уверенностью утверждать, что расщепление пропозиционального аргумента (с использованием предложно-падежной конструкции *как* + N_{Nom}) в предложениях с глаголом *удивлять/удивить* в современном русском языке является документированным явлением, и в этом отношении валентность данного глагол соответствует большинству единиц класса *verba sentiendi*.

ЗАКЛЮЧЕНИЕ

Корпусные данные исключительно важны для современных лингвистических исследований, как в области лингвистики языка, так и лингвистики речи. В синтаксических исследованиях обращение к корпусу позволяет верифицировать абстрактные синтаксические модели, документировать синтаксические связи, а также лексические реализации отдельных синтаксических позиций. Используя данный инструмент, мы устанавливаем ча-

стоту употребления тех или иных синтаксических форм, а тем самым и продуктивность синтаксических моделей.

Практика исследований показывает, что у корпусного анализа есть и определенные ограничения. Корпус оказывается недостаточно репрезентативным в случае менее частотных явлений и особенно тех, которые маркированы со стилистической точки зрения. Объективность корпусных данных повышается, если дополнительно используются другие источники материала: интроспекция (наблюдение, рефлексия), анкетирование, эксперимент, качественный и количественный анализ текстов, анализ словарных дефиниций.

Обращение к интроспекции (а также к языковой интуиции) может показаться необоснованным, особенно с учетом факта, что интернет-корпусы содержат миллионы примеров употреблений слов, а среднестатистический пользователь языка знает их только несколько десятков тысяч. Объяснение этого факта кроется в том, что корпусная разметка обычно касается характеристик единиц, и только в случае синтаксических корпусов возможен поиск по их отношениям, но этот поиск, следует признать, еще не совершенен. Языковая же компетенция включает не только единицы и их свойства, но также их системные отношения, включая динамические процессы и возможные состояния системы (в том числе и так называемые потенциализмы).

Из опыта корпусных исследований вытекает общий вывод, что нельзя отождествлять естественный язык ни с каким компендиумом форм его проявления. Так же, как, согласно Протагору, человек — это мера всех вещей, он также является мерой идентичности языковой системы. Понятие языкового факта не ограничивается материальной формой единицы большего или меньшего формата — оно всегда предполагает более или менее социализированное и, в принципе, латентное по способу существования знание носителей языка о контекстах употребления единиц. В связи с этим к идеологии популярного сегодня панэмпиризма (см. Андрияхина 1992: 18; Dębowski 2007: 120) следует относиться с осторожностью.

REFERENCES

Andryukhina, Lyudmila M. *Stil' nauki: kul'turno-istoricheskaya priroda*. Yekaterinburg: Izdatel'stvo Yekaterinburgskogo universiteta, 1992 [Андрияхина, Людмила

- М. *Стиль науки: культурно-историческая природа*. Екатеринбург: Издательство Екатеринбургского университета, 1992.
- Biber, Douglas, and Fitzmaurice, Susan, and Reppen, Randi. Eds.. *Using Corpora to Explore Linguistic Variation*. Amsterdam: Benjamins, 2002.
- Dębowski, Józef. "Wolność nauki a zasada bezzłozeniowości. Między «epistemologicznym anarchizmem» a «epistemologicznym restrykcyjizmem»." *Wolność w epoce poszukiwań*. Szulakiewicz, Marek, and Karpus. Zbigniew Eds. Toruń: Wydawnictwo UMK, 2007: 109–122.
- Gries, Stefan G., "What is Corpus Linguistics?" *Language and Linguistics Compass* 2009, no. 3: 1–17.
- Karolak, Stanisław. "Składnia wyrażen predykatywnych." *Gramatyka współczesnego języka polskiego. Składnia*. Topolińska, Zuzana. Ed. Warszawa: PWN, 1984: 11–212.
- Karolak, Stanisław. *Podstawowe struktury składniowe języka polskiego*. Warszawa: Instytut Sławistyki PAN, 2002.
- Kiklewicz, Aleksander, and Korytkowska, Małgorzata, and Mazurkiewicz-Sułkowska, Julia et al. *Zintegrowany opis semantyczno-syntaktyczny czasowników bułgarskich, polskich i rosyjskich (verba cogitandi i verba sentiendi)*. Część I. Warszawa: Instytut Sławistyki PAN & Fundacja Sławistyczna, 2019.
- Kiklewicz, Aleksander, and Korytkowska, Małgorzata. *Podstawowe struktury zdaniowe współczesnych języków słowiańskich: białoruski, bułgarski, polski*. Olsztyn: CBEW, 2010.
- Kiklewicz, Aleksander. "Sintaksicheskiye kharakteristiki russkikh i pol'skikh internet-forumov (na materiale prostykh i slozhnykh predlozheniy s mental'nymi predikatami)." *Interaktion von Internet und Stilistik, Internet und Stil*. Branko Tosović, and Arno Wonisch Eds. Graz: Universität Graz, 2016: 93–110 [Kiklewicz, Aleksander. "Синтаксические характеристики русских и польских интернет-форумов (на материале простых и сложных предложений с ментальными предикатами)." *Interaktion von Internet und Stilistik, Internet und Stil*. Branko Tosović, and Arno Wonisch. Eds. Graz: Universität Graz, 2016: 93–110].
- Korytkowska, Małgorzata, and Małdziejewa, Wiara. *Od zdania złożonego do zdania pojedynczego. Nominalizacja argumentu propozycjonalnego w języku polskim i bułgarskim*. Toruń: Wydawnictwo UMK, 2002.
- Kotin, Michail L. "Möglichkeiten und Grenzen der korpusbasierten Grammatikforschung." *Biuletyn Polskiego Towarzystwa Językoznawczego* 2018, no. LXXIV: 61–70.
- Marzo, Stefania, Heylen, Kris, Sutter Gert De. "Developments in Corpus-based Contrastive Linguistics." *Corpus Studies in Contrastive Linguistics*. Marzo, Stefania, and Heylen, Kris, and Sutter Gert De. Eds. Amsterdam: Benjamins, 2012: 1–7.
- Meurers, Detmar W. "On the use of electronic corpora for theoretical linguistics: Case studies from the syntax of German." *Lingua* 2005, no. 115/11: 1619–1639.
- Münzberg, Franziska. "Korpusrecherche in der Dudenredaktion: Ein Werkstattbericht." *Grammatik und Korpora 2009. Dritte Internationale Konferenz. Mannheim, 22.–24.9.2009*. Konopka, M., and Kubczak, J. and Jacqueline, C. Mair et al. Eds. Tübingen: Narr Verlag, 2011: 181–197.
- Mustajoki, Arto. "The Integrum Database as a Powerful Tool in Research on Contemporary Russian." *Integrum: tochnye metody i gumanitarnye nauki*. Galina Nikiporets-Takigava. Ed. Moskva: Letniy sad, 2006: 50–75.

- Piątkowski, Łukasz. "Zastosowanie analiz korpusowych w badaniach leksyko-gramatycznych. Perspektywa porównawcza niemiecko-polska." *Zeszyty Naukowe Towarzystwa Doktorantów UJ Nauki Humanistyczne* 2018, no. 21 (2): 131–150.
- Plungyan, Vladimir A. "Zachemmy delayem natsional'nyu korpus russkogo yazyka?" *Otechestvennyye zapiski* 2005, no. 2: 296–308 [Плунгян, Владимир А. "Зачем мы делаем национальный корпус русского языка?" *Отечественные записки* 2005, no. 2: 296–308].
- Polański, Kazimierz. *Słownik syntaktyczno-generatywny czasowników polskich. 1/A-M*. Wrocław etc.: Ossolineum, 1980.
- Urmantsev, Yuriy A. "Nachala obshchey teorii sistem." *Sistemnyy analiz i nauchnoye znaniye*. Dmitriy P. Gorskiy. Ed. Moskva: Nauka, 1978. 7–41 [Урманцев, Юрий А. "Начала общей теории систем." *Системный анализ и научное знание*. Дмитрий П. Горский. Ed. Москва: Наука, 1978. 7–41].
- Yevgen'yeva, Anastasiya P. *Slovar' russkogo yazyka*. T. 4. Moskva: Nauka, 1984 [Евгеньева, Анастасия П. *Словарь русского языка*. Т. 4. Москва: Наука, 1984].



MACIEJ LABOCHA

 ORCID: <https://orcid.org/0000-0002-8763-2187>

Uniwersytet Śląski w Katowicach

KORPUSY WOBEC POSZUKIWANIA OPTYMALNYCH KOLOKACJI DO CELÓW LEKSYKOGRAFII PRZEKŁADOWEJ

USING CORPORA IN SEARCH OF OPTIMAL COLLOCATIONS FOR TRANSLATION DICTIONARIES

The purpose of this article is to verify corpus findings for first fifty confirmations of Russian lexemes прожить and пережить (seen as nonmetaphorical equivalents of the Polish entry unit przeżyć) and identification of collocations that would be useful in lexicographical and didactic aspects. The article presents didactic and lexicographical research context and shows previously identified language translations pairs with definitions of original entries and collocational chains, both in terms of collocational schemes and chains of typical collocants. As well as this, the author quotes corpus confirmations for first fifty findings in the basic and press corpora respectively within the National Russian Corpus. In addition, the article includes potential complementary Russian equivalents for the Polish entry unit przeżyć.

WSTĘP

Poza wszelkimi wątpliwościami wydaje się obecnie pozostawać kwestia korelacji badań korpusowych z metaleksykografią i docelowo leksykografią jako praktyką układania słowników. Jednocześnie można odnieść wrażenie, że analiza korpusowa, mająca na celu opis lingwistyczny, i – z drugiej strony – analiza korpusowa, której celem jest odpowiedni opis jednostek języka do celów leksykograficznych to dwa zbiory o części wspólnej, podążające jednak w nieco innym kierunku. O ile w pierwszym przypadku możemy się zamknąć w analizach wybranych zjawisk językowych, to w drugim – analizy takie co do zasady powinny zmierzać do precyzyjnego opisu poszczególnych jednostek języka w sensie holistycznym. To jednak nastęcza pewnych trudności, wynikających zarówno z narzędzi korpusowych, jak ogólnometodologicznych.

W artykule niniejszym postaramy się w związku z tym odpowiedzieć na następujące pytania:

1) Czy opieranie się na inicjalnych wynikach korpusowych pozwala na wychwycenie wystarczającej liczby kolokacji, które można by wykorzystać w celach leksykograficznych?

2) Czy istnieje podejrzenie, że wyszukiwana słowoforma może ograniczać zakres otrzymywanych wyników?

3) Czy kolokacje *takie-jakie-są*, otrzymane w inicjalnych wynikach korpusowych są dobrym źródłem materiału ilustracyjnego dla słowników przekładowych?

4) Czy istnieją jakieś efekty uboczne poszukiwań korpusowych kolokacji w stosunku do wyodrębnionych uprzednio znaczeń i ustalonych par przekładowych?

Aby spróbować odpowiedzieć na te pytania przedstawimy rezultaty badania korpusowego opartego na inicjalnych „próbkach” (50 pierwszych wyników dla słowoform *прожить* oraz *пережить* jako bezpośrednich niemetaforycznych ekwiwalentów polskiej jednostki *przeżyć*) z dwóch podkorpusów Narodowego Korpusu Języka Rosyjskiego (<https://ruscorpora.ru/>), a mianowicie Korpusu podstawowego (Основной корпус) oraz Korpusu prasowego (Газетный корпус) (w artykule pełne nazwy tych subkorpusów używane będą zamiennie z abrewiaturami KPodst NKJR i KPras NKJR). Wybór obu podkorpusów podyktowany jest ich pierwotnie pisemną formą oraz potencjalnym realizowaniem stylu standardowego, a więc co do zasady właśnie tej płaszczyzny systemu językowego, która pozbawiona jest elementów stylów, niebędących realizacją szeroko pojmowanego języka literackiego czy narodowego, a co za tym idzie, odznaczającego się — jak można przypuszczać — wysokim stopniem normatywności.

Ponieważ niniejszy artykuł jest w istocie rozszerzeniem rozważań podejmowanych w publikacji *Glottodydaktyczne aspekty opisu słownikowego kolokacji*¹, a różni się od tego tekstu postawą metodologiczną (przede wszystkim orientacją na korpusy i pracę z nimi) — zostanie również zarysowany szerszy kontekst aktualnego badania wynikający z badań poprzednich.

KONTEKST BADAWCZY

O ile we wspomnianej publikacji głównym celem było przedstawienie glottodydaktycznego punktu widzenia na ekscerpcję oraz rejestrowa-

¹ M. Labocha, *Glottodydaktyczne aspekty opisu słownikowego kolokacji*, „Roczniki Humanistyczne” 2021, nr 10 (69), s. 121–138.

nie związków wyrazowych w słownikach przekładowych, to w niniejszym artykule chcielibyśmy nieco pogłębić opis korpusowych możliwości odpowiedniej ekscerpcji kolokantów wraz z ich charakterystyką, wynikającą z korpusowego opracowania materiału tekstowego, a także lingwistycznej recepcji badacza. We wspomnianym artykule uznaliśmy m.in., że granica pomiędzy leksykografią przekładową *sensu largo* oraz leksykografią dydaktyczną zaciera się², a także, że leksykografię przekładową powinny definiować możliwości percepcyjne użytkownika, natomiast sam opis słownikowy powinna charakteryzować tzw. elegancja, rozumiana jako opis „wyczerpujący i dokładny (w stopniu adekwatnym do potrzeb i oczekiwań odbiorcy), [...] zwięzły i prosty”³. Ową „elegancję”, choć nie wprost, zestawiliśmy również z refleksjami nie czysto lingwistycznymi, lecz pozornie dalekimi od tego, co można by nazwać klasyczną lingwistyką, a ściślej rzecz ujmując – z refleksją neurobiologiczną, mówiącą o tym, jak uczy się mózg⁴, co można w zupełności uznać za potwierdzenie pewnych progresywnych i holistycznych zarazem konkluzji wobec leksykografii przekładowej, która dotychczas raczej nie miała dobrej opinii wśród badaczy⁵. Badając kolokacje rosyjskich leksemów *прожить* i *пережить*, będących podstawowymi nie-metaforycznymi ekwiwalentami polskiego leksemu *przeżyć*, w oparciu o najprostszą metodę wyszukiwania w Narodowym Korpusie Języka Rosyjskiego (wyszukiwanie bez jakichkolwiek ram dla formy podstawowej tego czasownika), doszliśmy do wielu semantycznych modeli łączliwości wraz z szeregiem kolokantów. Owe szeregi (semantyczne i kolokacyjne) wyglądały następująco:

² Zob. S. Atkins, *Bilingual dictionaries: Past, Present and Future*, w: M. Gellerstam i in. (red.), *Euralex '96 Proceedings I*, Göteborg University, Göteborg 1996, s. 526; М.В. Влавацкая, «Users' needs» как требования к составлению англо-русского учебного комбинаторного словаря, „Вопросы лексикографии” 2013, nr 2 (4), s. 33–34.

³ M. Bańko, *Za mało czy dość? Dylemat leksykografa w dobie rewolucji informacyjnej (na przykładzie słowników frazeologicznych)*, w: G. Dziamska-Lenart, J. Liberek (red.), *Perspektywy współczesnej frazeologii polskiej. Między teorią a praktyką frazeograficzną*, Wydawnictwo UAM, Poznań 2013, s. 10–11.

⁴ M. Spitzer, *Jak uczy się mózg?* Wydawnictwo Naukowe PWN, Warszawa 2012, s. 63, 68.

⁵ Zob. opinie dotyczące słowników dwujęzycznych (przekładowych): J. Pieńkos, *Podstawy przekładoznawstwa. Od teorii do praktyki*. Kantor Wydawniczy Zakamycze, Kraków 2003, s. 174; J. Mędelska, J. Wawrzyńczyk, *Między oryginałem a przekładem. Rzecz o słownikach dwujęzycznych*. Wydawnictwo Szumacher, Kielce 1992, s. 9; T. Piotrowski, *Ekwiwalencja w słownikach dwujęzycznych*, w: W. Chlebda (red.), *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*, Wydawnictwo Uniwersytetu Opolskiego, Opole 2011, s. 50.

PRZEŻYĆ

Znaczenie 1) ‘żyjąc, przeżyć, przetrwać jakiś czas’

Ekwiwalent: *прожить*

Schemat łączliwości: *прожить* + ‘czas (określony lub nieokreślony)’

Typowe kolokanty: *жизнь / несколько лет / год / месяц / день* etc.

Znaczenie 2) ‘przetrzymać jakieś wydarzenia, wypadki, nie stracić życia’

Ekwiwalent: *пережить*

Schemat łączliwości: *пережить* + ‘wydarzenie zagrażające życiu’

Subschemat 1: *пережить* + ‘nazwa wydarzenia o charakterze konfliktu zbrojnego lub wydarzeń z niego wynikających’

Subschemat 2: *пережить* + ‘nazwa zdarzenia o charakterze nieszczęśliwego wypadku, zdarzenia losowego lub klęski żywiołowej’

Subschemat 3: *пережить* + ‘nazwa długotrwałych zjawisk klimatycznych nie sprzyjających podtrzymaniu życia’

Subschemat 4: *пережить* + ‘nazwa zdarzenia o charakterze medycznym’

Typowe kolokanty:

Subschemat 1: *пережить* революцию / бомбардировку / уличные беспорядки

Subschemat 2: *пережить* пожар

Subschemat 3: *пережить* зиму

Subschemat 4: *пережить* инфаркт / инсульт / воспаление лёгких etc.

Znaczenie 3) ‘zostać przy życiu dłużej od kogoś lub czegoś’

Ekwiwalent: *пережить*

Schemat łączliwości: *пережить* + ‘nazwa osoby’

Typowe kolokanty: *мужа / жену / детей / сына / дочь*

Znaczenie 4) ‘spędzić [żyjąc] jakiś czas w jakimś miejscu’

Ekwiwalent: *прожить*

Schemat łączliwości: *прожить* + в + ‘toponim’ + ‘czas określony lub nieokreślony’

Typowe kolokanty: в Германии 21 год / во Франции полжизни / в Париже двадцать лет / всю жизнь в городе (в провинции)

Znaczenie 5) ‘spędzić [żyjąc] jakiś czas w relacji z kimś’

Ekwiwalent: *прожить*

Schemat łączliwości

Subschemat 1: *прожить* + вместе + ‘czas (określony lub nieokreślony)’

Subschemat 2: *прожить* + с + nazwa osobowa w N. + ‘czas (określony lub nieokreślony)’

Subschemat 3: *прожить* + в + nazwa relacji w. Msc. + ‘czas (określony lub nieokreślony)’

Typowe kolokanty:

Subschemat 1: вместе всю жизнь / четверть века / тридцать лет / 10 лет,

Subschemat 2: с отцом сорок лет / с женой одиннадцать лет / с Иолантой одиннадцать лет,

Subschemat 3: в браке не один год

Znaczenie 6) ‘poradzić sobie w codziennym życiu, mając do dyspozycji minimalne środki materialne’

Ekwiwalent: *прожить*

Schemat łączliwości: *прожить* + на + ‘określona lub nieokreślona kwota pieniężna / nazwa świadczenia pieniężnego’ w formie B.

Typowe kolokanty: на 1137 руб. / на стипендию в 25 рублей / на литературные заработки / на одну зарплату / на пенсию / на пособие

Wspomniany aspekt glottodydaktyczny i wcześniejsze nasze rozważania warto jednak poszerzyć o kilka uwag oraz wniosków na temat wykorzystania słowników przekładowych/dwujęzycznych⁶ w kontekście glottodydaktycznym. Przede wszystkim należy podkreślić pewien paradoks słowników, w tym przekładowych. Ponownie porównajmy słowa badaczek napisane w odstępnie niemalże dwu dekad (symptomatyczny pozostaje tu fakt, że pierwszy z cytatów dotyczy słowników z językiem słoweńskim, drugi zaś — słowników z językiem niemieckim, a więc można byłoby wnioskować, że są to / powinny być niejako głosy w odrębnych tradycjach leksykograficznych):

Leksykografia dwujęzyczna, podobnie jak jednojęzyczna jest próbą ogarnięcia nieograniczonego oceanu słów. W każdym słowniku dwujęzycznym można bowiem znaleźć braki w zakresie ekwiwalencji, szczególnie tekstowej, bądź takie sytuacje użycia leksemu, które nie znalazły w nim odzwierciedlenia, niezależnie od tego, jak sumienny i dokładny jest leksykograf i czy ma on przygotowanie językoznawcze. [...] Ponieważ większość użytkowników traktuje słowniki dwujęzyczne jako naturalne słowniki do nauki języka obcego, powstaje pytanie, czy wobec wspomnianych ograniczeń mogą one dobrze pełnić tę funkcję?⁷

Tradycyjny, wydany w formie papierowej słownik jeszcze nie tak dawno stanowił podstawowe narzędzie wspierające działania językowe studenta uczącego się języka obcego. Obecnie jednak, w dobie cyfryzacji, wydaje się, że słowniki książkowe przegrywają w konfrontacji ze słownikiem elektronicznym. [...] Jednak o ile korzystanie ze słownika digitalnego ogranicza się zazwyczaj do konsultacji punktualnej, o tyle po słowniki drukowane sięga się obecnie znacznie częściej w momencie pojawienia się problemu nie tylko natury leksykalnej, lecz także ortograficznej, morfologicznej czy też składniowej⁸.

⁶ O tych dwu terminach zob. więcej w: Э. Бялэк, *Что такое переводной словарь? В поисках ответа*, „Вестник Московского университета. Сер. 22. Теория перевода” 2017, nr 4, s. 103–123.

⁷ B. Ostromięcka-Frażczak, *Słownik dwujęzyczny i jego rola w nauczaniu języka polskiego jako obcego*, „Acta Universitatis Lodziensis” 1998, nr 10, s. 389.

⁸ H. Kaczmarek, *Tradycyjny słownik książkowy w kształceniu językowym na poziomie akademickim*, „Prace Naukowe Akademii im. Jana Długosza w Częstochowie” 2016, t. XXV, s. 373–374.

Choć z jednej strony odnotowuje się fakt „braków [co najmniej – M.L.] w zakresie ekwiwalencji”, to z drugiej – wskazuje się na to, że słowniki dwujęzyczne traktowane są/były jako „naturalne słowniki do nauki języka obcego”, a słowniki w formie papierowej stanowiły „jeszcze nie tak dawno [...] podstawowe narzędzie wspierające działania językowe studenta uczącego się języka obcego” i były źródłem cieszącym się, mimo wszystko, pewnego rodzaju – metaforycznie rzecz ujmując – estymą czy autorytetem, tzn. są/były uznawane za źródła, do których użytkownicy skłonni są odwoływać się w charakterze źródła nadrzędnego, co sugerują słowa: „po słowniki drukowane sięga się obecnie znacznie częściej w momencie pojawienia się problemu nie tylko natury leksykalnej, lecz także ortograficznej, morfologicznej, czy też składniowej”. Mamy więc do czynienia z dotychczas chyba nadal nie rozwiązaną dychotomią „braki vs. referencjalne źródło nadrzędne”, przy czym dostrzeganie owych braków bardziej właściwe jest badaczom, a postrzeganie słowników jako referencjalnych źródeł nadrzędnych – niesłownikarzom, nielingwistom, nietłumaczom, albo adeptom każdej z tych „sztuk/dziedzin”, czyli (jeszcze) niespecjalistom w sensie najbardziej ogólnym.

Powołując się raz jeszcze na cytowane badaczki, należy zgodzić się z sformułowanymi przez nie wnioskami:

W dotychczasowej praktyce leksykograficznej różnica między słownikiem dwujęzycznym a pedagogicznym jednojęzycznym sprowadza się do tego, że słownik jednojęzyczny jest pomocny przy poprawnym użyciu wyrazów, a więc może być wykorzystywany do sprawdzenia poprawności już gotowego tekstu, dzięki zaś słownikowi dwujęzycznemu taki tekst może w ogóle zostać napisany. [...] Idealnym, oczywiście, byłby słownik, w którym byłyby połączone wspomniane wyżej funkcje⁹.

Budowanie złożonej świadomości leksykograficznej nie jest zadaniem łatwym i wymaga wielu zaplanowanych i ukierunkowanych działań. Wykorzystywanie słowników na zajęciach pozwoli studentom osiągnąć odpowiednią wiedzę na temat różnego typu słowników oraz na temat rodzaju informacji w nich zawartych. Wiedza ta powinna być poparta aktywnym treningiem praktycznym [...]. Konieczne jest także uwrażliwienie studentów na pewne niedoskonałości słowników dwujęzycznych [...]¹⁰.

Trudno jednak zgodzić się, że taki stan zawieszenia pomiędzy słownikami jednojęzycznymi i dwujęzycznymi jest czy powinien być

⁹ B. Ostromecka-Frączak, *Słownik dwujęzyczny i jego rola w nauczaniu...*, s. 393.

¹⁰ H. Kaczmarek, *Tradycyjny słownik książkowy...*, s. 383.

czymś naturalnym. Do lepszej jakości słowników dwujęzycznych — wychodzącej poza ograniczające pojęcie „tradycji leksykograficznej” — nie tylko można, ale i trzeba naukowo dążyć. Można to jednak zrobić jedynie wtedy, gdy m.in. metody opracowania materiałów źródłowych, będących podstawą „układania” słowników będą szeroko i krytycznie (w tym autokrytycznie) dyskutowane.

CHARAKTERYSTYKA PODKORPUSÓW NKJR

We wstępie do niniejszego artykułu wspomnieliśmy, że w części analitycznej będziemy korzystać z dwu podkorpusów Narodowego Korpusu Języka Rosyjskiego. Są to: Korpus podstawowy (Основной корпус) oraz Korpus prasowy (Газетный корпус). Aby potwierdzić słuszność wyboru należy jednak scharakteryzować oba korpusy.

Korpus podstawowy (Основной корпус) jest domyślnym korpusem Narodowego Korpusu Języka Rosyjskiego i obejmuje teksty, odzwierciedlające rosyjski język literacki, a jego zasoby można podzielić na dwie grupy: 1) współczesne teksty pisemne (połowa XX w. — początek XXI w.) oraz 2) teksty wczesne (połowa XVIII w. — połowa XX w.). Domyślnie wyszukiwanie w obu tych grupach odbywa się jednocześnie. Rodzaje tekstów zawartych w grupie pierwszej to m.in.: różne gatunki współczesnej prozy, teksty należące do gatunku dramatu, teksty o charakterze wspomnieniowo-biograficznym, publicystyka w czasopiśmie i gazetach, krytyka literacka, wiadomości prasowe, teksty naukowe, popularnonaukowe oraz edukacyjne, teksty religijne i religijno-filozoficzne, teksty dotyczące zagadnień produkcyjno-technicznych, teksty o charakterze oficjalnym (biznesowe i prawne) oraz teksty dotyczące życia codziennego (w tym teksty nie przeznaczone do publikacji — dzienniki, korespondencja osobista etc.). W grupie drugiej są to również teksty rozlicznych gatunków (literatura piękna, teksty naukowe, korespondencja osobista, publicystyka etc.), przy czym odsetek tekstów literatury pięknej w stosunku do grupy pierwszej jest tu wyższy.

Korpus prasowy (Газетный корпус) jest natomiast korpusem mocno ograniczonym ramami czasowymi. Jest to korpus obejmujący teksty prasowe z pierwszej dekady XXI w., a zatem posiada on bardziej aktualny charakter. W jego zasoby wchodzi teksty takich źródeł drukowanych, jak „Известия”, „Советский спорт”, „Труд”,

„Комсомольская правда”, ale także teksty agencji publikujących w formacie elektronicznym – РИА „Новости”, РБК, „Новый регион”¹¹.

ANALIZA WYNIKÓW KORPUSOWYCH

Biorąc pod uwagę postawione we wstępie pytania, na które chcieliśmy uzyskać odpowiedź lub chociażby zarys odpowiedzi, podpowiadającej dalsze działania w zakresie optymalnej procedury poszukiwania odpowiednich kolokantów, wyszukiwanie zarówno w Korpusie podstawowym NKJR (KPodst NKJR), jak i Korpusie prasowym NKJR (KPras NKJR) prowadziliśmy dla jednostek wyjściowych w formie bezokolicznika, a mianowicie – *прожить* oraz *пережить*, jako podstawowych form ekwiwalentnych wobec polskiej jednostki *przeżyć*. Inicjalne wyniki (50 pierwszych pozycji w każdym z korpusów) obejmują więc konteksty wyłącznie z tymi formami. W dalszej części przedstawimy wspomniane wyniki w podziale na poszczególne, ustalone wcześniej, znaczenia.

1) Przeżyć ‘żyjąc, przebyć, przetrwać jakiś czas’ — *прожить*

Kolokacje dla znaczenia ‘żyjąc, przebyć, przetrwać jakiś czas’ w określonym przedziale objętościowym w obu korpusach są najliczniejsze. Nie jest to zaskoczenie, o ile przyjmiemy, że znaczenie to jest nadrzędne. Przypomnijmy ustalony wcześniej schemat kolokacyjny:

PRZEŻYĆ

Znaczenie 1) ‘żyjąc, przebyć, przetrwać jakiś czas’

Ekwiwalent: *прожить*

Schemat łączliwości: *прожить* + ‘czas (określony lub nieokreślony)

Typowe kolokanty: *жизнь / несколько лет / год / месяц / день* etc.

Zarówno w przypadku Korpusu podstawowego NKJR, jak i w przypadku Korpusu prasowego NKJR wśród inicjalnych 50 kontekstów przeważają kolokacje z kolokantem *жизнь*, przy czym struktura składniowa jest luźna pod względem szyku oraz kolokant *жизнь* niejednokrotnie występuje wraz z odpowiednią przydawką:

¹¹ <https://ruscorpora.ru/new/corpora-structure.html> (18.12.2021).

(KPodst NKJR) *прожить жизнь; жизнь прожить; прожить долгую счастливую жизнь; прожить всю жизнь; прожить вторую жизнь; прожить ещё одну жизнь; прожить тусклую, тяжёлую жизнь; прожить несчастливую жизнь;* (KPras NKJR) *всю жизнь прожить; прожить долгую жизнь; прожить долгую творческую жизнь.* Najbardziej jaskrawym potwierdzeniem słabej pozycji składniowej może być poświadczenie zaczerpnięte z KPras NKJR: *Первая — ценность человеческой **жизни** и ответственность перед мирозданием за то, чтобы ее **прожить**, несмотря на все ужасы, что происходят вокруг, царящий абсурд и произвол.* Ów czas (określony lub nieokreślony) w ustalonym schemacie łączliwości może z kolei być fragmentowany: (KPras NKJR) *прожить очень небольшую часть своей жизни.*

Nieliczne są natomiast kolokacje dla znaczenia ‘żyjąc, przebyć, przetrwać jakiś czas’ z kolokantami *год/лета, месяц, день* etc., por.: (KPodst NKJR) *прожить сто лет; прожить четыре-пять лет; прожить несколько месяцев; денька бы три, четыре прожить; прожить оставшиеся дни;* (KPras NKJR) *прожить не 41 год, а 82 года и даже больше; прожить максимум несколько месяцев; прожить каждый отведенный ей день; прожить всего несколько минут [после рождения].*

W przypadku KPras NKJR odnajdujemy również pojedyncze konstrukcje, które określają punkt referencyjny w określeniu długości życia, tzn. konstrukcje wykorzystujące stopień wyższy oznaczający przekroczenie pewnej granicy czasowej lub, wraz z partykułą *не*, oznaczający nie wyjście poza taką granicę: *прожить больше 50 лет; прожить не более полутора лет.* Jest to wariant syntaktyczno-łączliwościowy, który być może warto odnotować jako uzupełniający subschemat kolokacyjny.

2) Przeżyć ‘przetrzymać jakieś wydarzenia, wypadki, nie stracić życia’ — пережить

Liczba wyekscerpowanych kolokacji dla powyższego znaczenia jest niższa niż dla znaczenia pierwszego (‘żyjąc, przebyć, przetrwać jakiś czas’) — pamiętajmy, że pod uwagę brane było 50 pierwszych wyników z obu korpusów NKJR. Ustalony wcześniej schemat kolokacyjny zawierał kilka subschematów:

PRZEŻYĆ

Znaczenie 2) ‘przetrzymać jakieś wydarzenie, wypadki, nie stracić życia’

Ekwiwalent: *пережить*

Schemat łączliwości: *пережить* + ‘wydarzenie zagrażające życiu’

Subschemat 1: *пережить* + ‘nazwa wydarzenia o charakterze konfliktu zbrojnego lub wydarzeń z niego wynikających’

Subschemat 2: *пережить* + ‘nazwa zdarzenia o charakterze nieszczęśliwego wypadku, zdarzenia losowego lub klęski żywiołowej’

Subschemat 3: *пережить* + ‘nazwa długotrwałych zjawisk klimatycznych nie sprzyjających podtrzymaniu życia’

Subschemat 4: *пережить* + ‘nazwa zdarzenia o charakterze medycznym’

Typowe kolokanty:

Subschemat 1: *пережить* революцию / бомбардировку / уличные беспорядки,

Subschemat 2: *пережить* пожар

Subschemat 3: *пережить* зиму

Subschemat 4: *пережить* инфаркт / инсульт / воспаление лёгких etc.

Subschemat 1, a mianowicie *пережить* + ‘nazwa wydarzenia o charakterze konfliktu zbrojnego lub wydarzeń z niego wynikających’ na podstawie inicjalnych 50 wyników można zilustrować jedynie kilkoma przykładami z korpusu prasowego: (KPras NKJR) *пережить зомби-апокалипсис; пережить ужасы войн и терактов*; ale także *пережить последствия ядерной войны*. W ostatnim przypadku kolokacja jest przez nas doprowadzona do postaci zwartej z wyimka, w którym jednostka *ядерная война* jest ze względów składniowych zastąpiona w bezpośredniej pozycji kolokacyjnej zaimkiem *её*: *В песне Fallout Shelter Билли Чамберс рассказывает трогательную историю любви: начинается ядерная война, и отец успокаивает сына — все будет в порядке, ведь у нас есть свой бункер, в котором мы сможем пережить ее последствия*.

W przypadku subschematu 2 — *пережить* + ‘nazwa zdarzenia o charakterze nieszczęśliwego wypadku, zdarzenia losowego lub klęski żywiołowej’ udało nam się wyeksceperować zaledwie jedno poświadczenie kolokacji tego typu: (KPodst NKJR) *пережить [не один] голод*.

Subschemat 3, obejmujący składniki *пережить* + ‘nazwa długotrwałych zjawisk klimatycznych nie sprzyjających podtrzymaniu życia’ również reprezentowany jest w 50 inicjalnych wystąpieniach w obu korpusach łącznie jedynie kilkoma poświadczeniami: (KPodst NKJR): *пережить зиму; пережить зимние холода*; (KPras NKJR) *пережить [наступившую] зиму*.

Podobnie rzecz ma się z subschematem 4. o strukturze *пережить* + ‘nazwa zdarzenia o charakterze medycznym’. Poświadczenia korpusowe w 50 pierwszych wystąpieniach są równie nieliczne i odnajdujemy je jedynie w Korpusie prasowym: (KPras NKJR) *пережить процедуру*; *(не) пережить хирургического вмешательства*; *пережить несколько хирургических вмешательств*. W przypadku ostatniego przykładu można mieć wątpliwości, czy ekwiwalentem kontekstowym rosyjskiej jednostki *пережить* jest polski leksem *przeżyć* — wydaje się, że lepszym odpowiednikiem byłby tu polski wyraz *przeżyć*, a całe wyrażenie można by przetłumaczyć jako *przeżyć kilka ingerencji/zabiegów chirurgicznych*.

Wśród potencjalnych kolokacji, które można rozważać jako odnoszące się do tego znaczenia nadrzędnego, tj. ‘przetrzymać jakieś wydarzenia, wypadki, nie stracić życia’ odnajdujemy natomiast (KPras NKJR) *пережить 90-е и нулевые*. Trudno jednak jednoznacznie przyporządkować tę jednostkę do któregoś z określonych subschematów kolokacyjnych — być może warto zmodyfikować sformułowane wcześniej lub rozszerzyć ich spis o kolejny, np. *пережить* + ‘nazwa okresów, wydarzeń o charakterze społecznym lub ekonomicznym’.

3) Przeżyć ‘zostać przy życiu dłużej od kogoś lub czegoś’ — *пережить*

Wśród 50 inicjalnych wyników w Korpusie podstawowym NKJR oraz Korpusie prasowym NKJR odnaleźliśmy tylko po jednym wystąpieniu kolokacji potwierdzających wskazany poniższej schemat łączliwości dla znaczenia ‘zostać przy życiu dłużej od kogoś lub czegoś’:

PRZEŻYĆ

Znaczenie 3) ‘zostać przy życiu dłużej od kogoś lub czegoś’

Ekwiwalent: *пережить*

Schemat łączliwości: *пережить* + ‘nazwa osoby’

Typowe kolokanty: мужа / жену / детей / сына / дочь

Rejestrowane w tej próbce kolokacje to odpowiednio: (KPodst NKJR) [*не*] *пережить своего царя и благодателя*; (KPras NKJR) *пережить [не только] королеву*.

4) Przeżyć ‘spędzić [żyjąc] jakiś czas w jakimś miejscu’ — прожить

Schemat kolokacyjny dla znaczenia ‘spędzić [żyjąc] jakiś czas w jakimś miejscu’ potwierdzają jedynie dwa poświadczenia spośród pierwszych 50 wyników w Korpusie podstawowym NKJR. W Korpusie prasowym NKJR nie odnaleźliśmy natomiast dostatecznych poświadczeń. Ustalony uprzednio schemat kolokacyjny wyglądał następująco oraz zawierał następujące kolokanty:

Znaczenie 4) ‘spędzić [żyjąc] jakiś czas w jakimś miejscu’

Ekwiwalent: *прожить*

Schemat łączliwości: *прожить* + в + ‘toponim’ + ‘czas określony lub nieokreślony’

Typowe kolokanty: в Германии 21 год / во Франции полжизни / в Париже двадцать лет / всю жизнь в городе (в провинции)

Przykłady kolokacji wskazanych przez Korpus podstawowy NKJR to: *прожить 21 год в Германии; прожить в стране не меньше пяти лет*, przy czym drugie z poświadczeń nie posiada silnej pozycji składniowej, a rzeczownik *страна* jest w nim pierwotnie zamieniony zaimkiem *она* w formie Msc.: *Типы прозы увидеть в чужой стране труднее, для этого надо прожить в ней, наверное, не меньше пяти лет, однако Холдена Колфилда я встречал, и не раз, и в кампусе, и в городе, и на дорогах, и юношу Холдена, и мужчину Колфилда, и старика мистера Холдена Колфилда.*

W obu przypadkach można mieć jednak wątpliwości, czy za najbardziej naturalny ekwiwalent w języku polskim należy uważać leksem *przeżyć* — szczególnie w drugim, którego kontekst podpowiada użycie odpowiedników przekładowych *spędzić* lub *przemieszkać*.

5) Przeżyć ‘spędzić [żyjąc] jakiś czas w relacji z kimś’ — прожить

Schemat łączliwości dla znaczenia ‘spędzić [żyjąc] jakiś czas w relacji z kimś’, podobnie jak i w znaczeniu 2., zawiera kilka subschematów łączliwościowych:

Znaczenie 5) ‘spędzić [żyjąc] jakiś czas w relacji z kimś’

Ekwiwalent: *прожить*

Schemat łączliwości

Subschemat 1: *прожить* + вместе + ‘czas (określony lub nieokreślony)’

Subschemat 2: *прожить* + с + nazwa osobowa w N. + ‘czas (określony lub nieokreślony)’

Subschemat 3: *прожить* + в + nazwa relacji w. Msc. + ‘czas (określony lub nieokreślony)’

Typowe kolokanty:

Subschemat 1: вместе всю жизнь / четверть века / тридцать лет / 10 лет

Subschemat 2: с отцом сорок лет / с женой одиннадцать лет / с Иолантой одиннадцать лет

Subschemat 3: в браке не один год

Zarówno Korpus podstawowy NKJR, jak i Korpus prasowy NKJR w 50 pierwszych wynikach podają nieliczne poświadczenia związane z tym znaczeniem: (KPodst NKJR) *прожить вместе 10 лет*; (KPras NKJR) *прожить вместе [хоть] 40 лет*; *прожить с Уиллисом 13 лет*. Poświadczenia te odpowiadają jedynie dwóm z trzech ustalonych subschematów kolokacyjnych.

Korpus podstawowy NKJR rejestruje jednak kilka innych konstrukcji, które można by uznać za przykłady odmiennego subschematu kolokacyjnego, por.: *не сможем прожить и одного дня друг без друга*; *не могли прожить друг без друга ни часа; дня не могли прожить друг без друга*; *без вас часа прожить не может*. Subschemat ten można rozpisać następująco:

Subschemat 4: *не мочь/смочь* + *прожить* + (и/ни + ‘jednostka czasu’ w D.) + (без + ‘nazwa osobowa’ w D. / друг без друга)

Typowe kolokanty: и/ни дня/часа без кого

6) Przeżyć ‘poradzić sobie w codziennym życiu, mając do dyspozycji minimalne środki materialne’ — прожить

Liczba poświadczeń w obu badanych korpusach dla znaczenia ‘poradzić sobie w codziennym życiu, mając do dyspozycji minimalne środki materialne’ polskiej jednostki *преżyć* i jej bezpośrednio rosyjskiego odpowiednika *прожить* jest relatywnie wyższa niż w przypadku większości znaczeń nieprzenośnych. Nie jest to jednak liczba zadowalająca, choć trzeba odnotować fakt, że zarówno Korpus podstawowy NKJR, jak i Korpus prasowy NKJR rejestrują porównywalną liczbę kolokantów. Przypomnijmy raz jeszcze wcześniejsze ustalenia, dotyczące schematu łączliwości i typowych kolokantów:

Znaczenie 6) ‘poradzić sobie w codziennym życiu, mając do dyspozycji minimalne środki materialne’

Ekwiwalent: *прожить*

Schemat łączliwości: *прожить* + на + ‘określona lub nieokreślona kwota pieniężna / nazwa świadczenia pieniężnego’ w formie В.

Typowe kolokanty: на 1137 руб. / на стипендию в 25 рублей / на литературные заработки / на одну зарплату / на пенсию / на пособие

Korpus podstawowy NKJR notuje następujące szeregi kolokacyjne: (KPodst NKJR) *прожить на литературные заработки* (dwukrotne wystąpienie); *прожить на зарплату*; *прожить [только] на [маленькую] пенсию*; *прожить на 1137 руб. [в месяц]*. Równolegle można ponownie odnotować użycie ww. jednostki w kontekście kolokacyjnym strukturalnie (składniowo) luźnym: *Сумма сама по себе громадная, но на неё надо было прожить месяц вчетвером* — kolokant należący do klasy wyrazów oznaczających ‘kwotę pieniężną / nazwę świadczenia pieniężnego’ zastąpiony jest w bezpośredniej pozycji składniowej odpowiednim zaimkiem.

Korpus prasowy NKJR rejestruje zaś takie tekstowe przykłady użycia tej jednostki: (KPras NKJR) *прожить на оставшуюся сумму*; *прожить на вырученные деньги*; *прожить на 100 долларов*; *прожить на сто тысяч рублей [в месяц]*; *на пенсию сложно прожить*. Biorąc pod uwagę ostatnie z przytoczonych poświadczeń ponownie możemy skonstatować, iż struktura składniowa kolokacji bywa luźna, por. *прожить на пенсию сложно* vs. *на пенсию сложно прожить*. Ponadto owa ‘kwota pieniężna / nazwa świadczenia pieniężnego’, będąca istotą znaczenia kolokantu może przybrać — jak pokazuje badanie korpusowe — semantycznie formę ‘kwoty nieokreślonej’ w dosłownym tego określenia znaczeniu: *смогут ли они прожить на ту сумму, которая будет оставаться у них после оплаты кредита*.

WNIOSKI POBOCZNE

Analiza korpusowa, opierająca się na poszukiwaniu kolokacji w oparciu o słowoformy *прожить* oraz *пережить* prowadzi również do wielu wniosków pobocznych dotyczących odkrywania niejako mimochodem jednostek przekładowych dla omawianych leksemów. W badaniu korpusowym na potrzeby leksykograficzne jest to — co zupełnie jasne — rezultat godny odnotowania.

Poszukując kolokacji dla jednostki *прожить*, natrafiiliśmy na liczne konteksty, dla których ekwiwalentnymi jednostkami języka polskiego mogłyby być m.in.:

1) wyżyć

(KPodst NKJR):

Наши люди, конечно же, не терпеливые китайцы, но повторю ещё раз — на одной продаже нефти и газа нам не **прожить**.

(KPras NKJR):

Теоретически я могу и полгода не получать заработную плату, если жена не убьёт до тех пор, то **прожить** можно спокойно, даже без зарплаты», — сказал украинский политик.

2) wytrzymać

(KPodst NKJR):

Знаете, когда мы недавно поехали отдыхать, то перед путешествием дали себе твёрдое слово: во время поездки ни слова о работе — ни о музыке, ни о медицине. Но... мы смогли **прожить** так только пять дней.

(KPras NKJR):

При постоянной работе GPS гаджет может **прожить** 30 часов, что тоже внушительно для небольшого по размеру, но мощного по емкости аккумулятора.

3) żyć

(KPodst NKJR):

Как можешь ты так **прожить**, чтобы Он радовался на тебя, радовался, видя, какой ты человек, каким ты стал ради этой дружбы.

(KPras NKJR):

И ты кто? Абсолютно один. Никогда нельзя желать **прожить** 120 лет. Иначе столько же надо желать прожить всем друзьям и врагам.

Pośród poświadczeń tekstowych dla rosyjskiej jednostki **пережить** napotkaliśmy zaś następujące przykładowe konteksty determinujące odpowiednie potencjalnie ekwiwalentne jednostki języka polskiego:

1) przejść / przechodzić (co, przez co)

(KPodst NKJR):

Желаю Вашим деткам-подросткам не **пережить** того гавна, которое я увидела в жизни.

(KPras NKJR):

Нездоровые привычки и другие способы расслабиться помогали вам **пережить** трудное детство, но вы больше не ребенок.

2) przetrwać

(KPodst NKJR):

Хорошо, если стране удастся **пережить** нынешнюю «стабилизацию» и она не перейдёт в полную и окончательную деградацию.

(KPras NKJR):

Бывший президент Украины Петр Порошенко заявил, что Украина может не **пережить** еще один государственный переворот.

3) przeczekać

(KPodst NKJR):

Ибо тогда встанёт вопрос: сможет ли телевидение технически развиваться и сможет ли оно закупать хорошие фильмы или права на спортивные трансляции? Может быть, этот период надо **пережить**. А потом мы придём к тому, что существует в развитых странах.

(KPras NKJR):

Ученые Университета Отаго в Новой Зеландии определили самые безопасные места на Земле, где человечество может **пережить** глобальную эпидемию и возродить технологическую цивилизацию.

PODSUMOWANIE

Odpowiadając na pytania postawione we wstępie do niniejszego artykułu, możemy sformułować następujące odpowiedzi będące jednocześnie wskazówką dotyczącą opracowywania materiału korpusowego do celów leksykografii przekładowej. Otóż:

1) opieranie się na inicjalnych wynikach korpusowych pozwala na wychwycenie pewnych schematów kolokacyjnych, ale nie w liczbie, która mogłaby być zadowalająca analitycznie (naukowo), a także nie w liczbie i jakości, która byłaby wystarczająca do obrazowej i użytecznej reprezentacji schematów kolokacyjnych na potrzeby leksykografii przekładowej, o której chcielibyśmy myśleć w kontekście pojęcia *użyteczności*;

2) uzyskane wyniki korpusowe, a w szczególności niska liczebność kolokacji dla niektórych ze znaczeń, pozwalają domniemywać, że ograniczenie się tylko do jednej słowoformy (w przypadku czasowników — forma bezokolicznika, formy osobowe czasu teraźniejszego (przyszłego prostego) / przeszłego czy form imiesłowowych) może potencjalnie wpływać na rezultat poszukiwań (w kontekście niniej-

szego badania problem ten pozostaje otwarty);

3) kolokacje *takie-jakie-są*, tzn. literalne wyimki z korpusu mogą lecz nie muszą być dobrym źródłem materiału ilustracyjnego na potrzeby leksykografii przekładowej (por. np. jednostki *пережить царя*; *пережить королеву* dla znaczenia 3. ‘zostać przy życiu dłużej od kogoś lub czegoś’ — jednostki te wydają się oddalone od przeciętnej recepcji podobnych zjawisk przez statystycznego użytkownika słownika przekładowego, nie tworzą też łatwego w odbiorze i czytelnego szeregu kolokacyjnego);

4) efektem pobocznym poszukiwań kolokacji korpusowych w stosunku do wyodrębnionych uprzednio znaczeń i ustalonych par przekładowych jest wyselekcjonowanie kolejnych par potencjalnych ekwiwalentów, które należałoby uwzględnić w filiacji (podziale na znaczenia/użycia) artykułu hasłowego wraz z odpowiednią semantyzacją (definicją zgodną z odpowiednim użyciem czy wskazówkami semantycznymi, dotyczącymi użycia poszczególnych jednostek) oraz wskazaniem — jeśli to możliwe — odpowiednich schematów kolokacyjnych.

W powyższym badaniu napotkaliśmy ponadto wiele poświadczeń korpusowych dla obu jednostek rosyjskich, co do których ze względu na kontekst mieliśmy wątpliwości natury definicyjnej (semantycznej), co mogłoby przekładać się — rzecz jasna — na realizację filiacji (przy założeniu *jednostka* vs. *jednostka*, a nie *jednostka* vs. *szereg jednostek*) w potencjalnie projektowanym słowniku przekładowym, a w wyżej wymienionym badaniu — na nieuwzględnienie kolokacji w korelacji z konkretnym znaczeniem. Wydaje się to (niestety) naturalnym rezultatem trudności w ustalaniu korelacji (*pojedyncza jednostka wejściowa* — (*pojedyncze*) *znaczenie* — (*pojedyncza*) *jednostka wyjściowa* — *kolokacje*, przy czym należy pamiętać, że poszukiwanie odpowiednich kolokantów na potrzeby leksykograficzne, uwzględniające pojęcie „użyteczności słownika” wymaga obowiązkowo uwzględnienia tych czterech korelatów łącznie. W przypadku ograniczonych danych niezastąpiona może okazać się zatem praktyka i intuicja badawcza, która pozwala rozwinąć i w dalszej kolejności uzupełnić lub zweryfikować otrzymane informacje w oparciu o metodę naukową w uzupełniających badaniach słownikowo-korpusowych.

REFERENCES

- Atkins, Sue. "Bilingual dictionaries: Past, Present and Future." *Euralex '96 Proceedings I*. Gellerstam, Martin et al. Eds. Göteborg: Göteborg University, Department of Swedish, 1996: 515–546.
- Bańko, Mirosław. "Za mało czy dość? Dylemat leksykografa w dobie rewolucji informacyjnej (na przykładzie słowników frazeologicznych)." *Perspektywy współczesnej frazeologii polskiej. Między teorią a praktyką frazeograficzną*. Dziamska-Lenart, Gabriela and Liberek, Jarosław. Eds. Poznań: Wydawnictwo Naukowe UAM, 2013: 9–16.
- Byalek, Eva, "Chto takoye perevodnoy slovar' V poiskakh otveta." *Vestnik Moskovskogo universiteta. Ser. 22. Teoriya perevoda*, 2017, no. 4: 103–123 [Бялэк, Эва, "Что такое переводной словарь? В поисках ответа." *Вестник Московского университета. Сер. 22. Теория перевода*, 2017, no. 4: 103–123.
- Kaczmarek, Hanna. "Tradycyjny słownik książkowy w kształceniu językowym na poziomie akademickim." *Prace Naukowe Akademii im. J. Długosza w Częstochowie*, XXV, 2016: 373–384.
- Labocha, Maciej, "Glottodydaktyczne aspekty opisu słownikowego kolokacji." *Roczniki Humanistyczne*, 2021, no. 10 (69): 121–138.
- Mędelaska, Jolanta, Wawrzyńczyk, Jan. *Między oryginałem a przekładem. Rzecz o słownikach dwujęzycznych*. Kielce: Wydawnictwo Szumacher, 1992.
- Ostromęcka-Frączak, Bożena. "Słownik dwujęzyczny i jego rola w nauczaniu języka polskiego jako obcego." *Acta Universitatis Lodzianensis*, 1998, no. 10: 398–393.
- Pieńkos, Jerzy. *Podstawy przekładoznawstwa. Od teorii do praktyki*. Kraków: Kantor Wydawniczy Zakamycze, 2003.
- Piotrowski, Tadeusz. "Ekwiwalencja w słownikach dwujęzycznych." *Na tropach translatorów. W poszukiwaniu odpowiedników przekładowych*. Chlebda, Wojciech. Ed. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2011. 45–69.
- Spitzer, Manfred. *Jak uczy się mózg?* Warszawa: Wydawnictwo Naukowe PWN, 2012.
- Vlavatskaya, M.V. "«Users' needs» kak trebovaniya k sostavleniyu anglo-russkogo uchebnogo kombinatornogo slovarya." *Voprosy leksikografii*, 2013, no. 2 (4): 32–40 [Влавацкая, М.В. "«Users' needs» как требования к составлению англо-русского учебного комбинаторного словаря." *Вопросы лексикографии*, 2013, no. 2 (4): 32–40].



MARIA MOCARZ-KLEINDIENST

 ORCID: <http://orcid.org/0000-0002-2205-5470>

Katolicki Uniwersytet Lubelski Jana Pawła II, Lublin

KORPUSY JĘZYKOWE W POLSKO-ROSYSKIM PROJEKCIE LEKSYKOGRAFICZNYM *ARCHITEKTURA I URBANISTYKA*

LANGUAGE CORPORA IN A POLISH-RUSSIAN LEXICOGRAPHICAL PROJECT *ARCHITECTURE AND URBAN PLANNING*

The article describes the usefulness of the language corpus resources and the corpus platform during the creation of the Polish-Russian dictionary „Architecture and urban planning.” The following tools were used: National Corpus of Polish Language, National Corpus of Russian Language, Polish-Russian Corpus of the University of Warsaw and Sketch Engine platform. It has been shown that the implementation of resources proceeds with varying intensity at different stages of dictionary development. Language corpora (together with available tools) are most helpful at the first stages, i.e. building dictionary entries (basic entries and their derivatives). To a much lesser extent, language corpora are used in the process of searching for Russian equivalents due to the relatively poor collection of specialised and even applied texts available in parallel corpora. Often the results obtained have only a probing value. It is important to combine data obtained from different corpus resources in order to obtain information relevant to the content of the headword article of the designed dictionary.

1. WPROWADZENIE

Postępujący proces cyfryzacji, możliwy dzięki rozwojowi technologii informatycznych, otworzył nowe perspektywy dla badań naukowych, w tym językoznawczych¹. Wymiernym rezultatem rewolucji cyfrowej są m.in. elektroniczne korpusy językowe, które dzięki zintensyfikowanym pracom zespołów badawczych na całym świecie zdążyły wykształcić różne odmiany: jednojęzyczne, dwujęzyczne, wielojęzyczne, zrównoważone oraz równoległe. Korpusy językowe ze swoim utylitar-

¹ G. Vetulani, *Problemy i korzyści wynikające z automatycznego przetwarzania korpusów — na przykładzie badań z zakresu predykcji rzeczownikowej w języku polskim*, „Roczniki Humanistyczne” 2013, LXI, z. 8, s.15.

nym charakterem, dzięki zgromadzonym w nich zasobom językowym z tekstowymi poświadczeniami, frekwencjami występowania i anotowanymi danymi metajęzykowymi, oraz dysponujące narzędziami ich automatycznego wyszukiwania i przetwarzania stały się podstawą rozwoju lingwistyki korpusowej². Wprowadziły także nową jakość badań do istniejących subdyscyplin językoznawstwa, m.in. translatoryki, glottodydaktyki, socjolingwistyki czy leksykografii³. W ramach tej ostatniej badacze dostrzegają i wykorzystują zasoby korpusowe jako źródło materiałowe w nowych projektach leksykograficznych⁴.

2. CEL, PRZEDMIOT I METODA BADAŃ

Przekonanie o potencjale leksykograficznym, a dokładniej hasłotwórczym zasobów korpusowych stało się punktem wyjścia w niniejszych badaniach. Mają one charakter empiryczny, aplikacyjny. Ich

² Zob.: B. Lewandowska-Tomaszczyk, *Podstawy językoznawstwa korpusowego*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2005.

³ Szerzej na temat możliwości wykorzystania korpusów językowych m.in. w pracach: P. Baker, *Sociolinguistics and Corpus Linguistics*, Edinburgh University Press, Edinburgh 2010; M.A.K. Halliday, W. Teubert, C. Yallop, A. Cermáková, *Lexicology and Corpus Linguistics. An Introduction*, Continuum, Londyn 2004; A. Kilgarriff, V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, V. Suchomel, *The Sketch Engine: Ten Years On*, „Lexicography” 2014, nr 1, s. 7 i n.; Ł. Grabowski, *Statystyka, programowanie i co jeszcze? O wyzwaniach współczesnego językoznawstwa*, w: W. Mokijenko, J. Tarsa (red.), *Język i pamięć. Księga jubileuszowa dedykowana Panu Profesorowi Wojciechowi Chlebdzie z okazji 70. urodzin*, Wydawnictwo Uniwersytetu Opolskiego, Opole 2020, s. 257–271.

⁴ Zob. np.: D. Bralewski, *Od przekładu do słownika. Korpus równoległy w redakcji słowników tłumaczeniowych*, Oficyna Wydawnicza LEKSEM, Łask 2012; A. Charciarek, *Korpus równoległy InterCorp w leksykografii przekładowej – możliwości i ograniczenia*, „Roczniki Humanistyczne” 2019, LXVII, z. 7; Д.О. Добровольский, *Корпусы текстов и двуязычная фразеография*, „Вестник Новосибирского государственного педагогического университета” 2015, nr 5(27); G. Faab, *Lexicography and Corpus Linguistics*, w: P.A. Fuertes-Olivera (red.), *The Routledge Handbook of Lexicography*, Routledge, Oxon 2017; Ł. Grabowski, *Korpusy dwu- i wielojęzyczne w służbie tłumacza, leksykografa i badacza: poszukiwanie ekwiwalentów przekładowych w świetle hipotez dotyczących istnienia uniwersaliów tłumaczeniowych*, w: W. Chlebda (red.), *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*, Wydawnictwo Uniwersytetu Opolskiego, Opole 2011; P. Hanks, *The corpus revolution in lexicography*, „International Journal of Lexicography” 2012, nr 25(4), s. 399 i n.

przedmiotem będzie prezentacja rzeczywistej przydatności korpusów językowych na poszczególnych etapach tworzenia konkretnego dwujęzycznego polsko-rosyjskiego projektu leksykograficznego z zakresu architektury i urbanistyki. Podjęte badania mają podwójne umocowanie w następujących po sobie działaniach praktycznych. Po pierwsze, sam pomysł opracowania słownika dwujęzycznego w kierunku polsko-rosyjskim narodził się w trakcie prac nad tłumaczeniem tekstów specjalistycznych na zlecenie Lubelskiego Wojewódzkiego Konserwatora Zabytków. Teksty były silnie nasycone terminologią głównie z zakresu architektury, w mniejszym stopniu urbanistyki. Zdecydowana większość terminów nie jest notowana w dostępnych dwujęzycznych ogólnych i specjalistycznych słownikach polsko-rosyjskich. Proces poszukiwań optymalnych odpowiedników przekładowych, nienotowanych w słownikach oznaczał czasochłonne „tropienie” potencjalnych ekwiwalentów w licznych, sprofilowanych dziedzinowo, ale w żaden sposób fizycznie niezespoleonych ze sobą tzw. tekstach paralelnych. Po drugie, już w trakcie realizacji autorskiego projektu pomysłu leksykograficznego uznano za zasadne, a nawet konieczne odwołanie się do dostępnych zasobów korpusowych. Jest to praktyka coraz częściej stosowana, zasadniczo wręcz wskazana w tego typu pracach. Jej potwierdzeniem są słowniki opracowane wyłącznie na podstawie elektronicznych danych korpusowych. Za przykład może posłużyć *Wielki słownik języka polskiego PAN* powstały na bazie korpusu Instytutu Podstaw Informatyki PAN, korpusu PWN oraz kilku wersji Narodowego Korpusu Języka Polskiego⁵. Narodowy Korpus Języka Rosyjskiego z kolei stał się bazą kilku tzw. słowników eksperymentalnych, wśród nich można wymienić na przykład *Словарь русской идиоматики. Сочетания слов со значением высокой степени* autorstwa G. I. Kustowej⁶. Te działania uzasadniają aplikacyjny charakter niniejszego studium. Od razu należy jednak zaznaczyć, że obiekt projektu leksykograficznego, jakim są terminy z zakresu architektury i urbanistyki, pozwala na zastosowanie metody analizy materiałów źródłowych jedynie wspomaganą korpusami (tzw. *corpus-assisted analysis*⁷), nie zaś opartej wyłącznie na ich zasobach. Dostępne w zbiorach korpusów językowych kolekcje tekstów specjalistycznych są zbyt skromne (jeśli w ogóle są obecne), by

⁵ P. Żmigrodzki (i in.) *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*, Instytut Języka Polskiego PAN, Kraków 2018, s. 19–20.

⁶ Słownik dostępny jest pod adresem: <http://dict.ruslang.ru/magn.php>.

⁷ P. Baker, *Sociolinguistics...*, s. 8.

mogły być wystarczające dla ekstrakcji materiału relewantnego dla dwujęzycznego (a nawet jednojęzycznego) słownika specjalistycznego. Niezbędne są inne źródła, w tym konsultacje ze specjalistami z zakresu historii sztuki. Ważne jest doprecyzowanie, że w niniejszych badaniach na plan pierwszy zostanie wysunięty aspekt realnych możliwości wykorzystania zasobów korpusowych, mimo istniejących ograniczeń, na które często zwracają uwagę badacze⁸. Świadomość istniejących ograniczeń nie wyklucza wykorzystania tego zasobu wiedzy językowej, sugeruje jedynie ostrożność, zachowawczość, czujność językową wobec pozyskiwanych informacji korpusowych dla celów leksykograficznych i oczywiście uzasadnia przyjętą metodę *corpus-assisted analysis*.

3. CHARAKTERYSTYKA PROJEKTU ARCHITEKTURA I URBANISTYKA

Specjalistyczny projekt leksykograficzny, w procesie powstawania którego podjęto próbę implementacji dostępnych zasobów i narzędzi korpusowych, ma charakter objaśniająco-przekładowy (tj. służący celom przekładowym). To oznacza, że hasła wyjściowe: podstawowe oraz podrzędne (tworzące razem układ gniazdowy) są opatrzone krótkim opisem w języku polskim, dopiero po nim pojawiają się odpowiedniki rosyjskie. Opis zawiera definicję encyklopedyczną objaśniającą znaczenie danego terminu, niekiedy jest ona wzbogacona krótką informacją na temat kontekstu historyczno-kulturowego doniosłego dla powstania danego terminu. Opis uzupełnia informacja w postaci przykładu obiektu, dla którego dane cechy, zjawiska, style lub detale są typowe, niekiedy nawet kluczowe. Słownictwo, które jest obiektem opisu i dwujęzycznego zestawienia, jest skupione wokół tematyki architektury i urbanistyki jako wytworów kultury, której immanentną cechą jest kumulowanie zarówno zjawisk uniwersalnych, wspólnych dla szerszego (europejskiego) kręgu kulturowego, zjawisk adaptowanych przez kultury narodowe geograficznie sytuujące się w takim szerszym kręgu, jak również zjawisk reprezentatywnych wyłącznie dla uwarunkowanej historycznie i geograficznie kultury narodowej. Na etapie ustalania koncepcji projektu wyznaczono polsko-rosyjski kierunek tworzenia par odpowiedności przekładowych. Wraz z nim

⁸ Por. np. A. Charciarek, *Korpus równoległy...*, s. 83 i in.; Ł. Grabowski, *Korpusy...*, s. 108 i n.; A. Kiklewicz, *Korpus internetowy jako źródło informacji lingwistycznej: ograniczenia*, „Prace Językoznawcze” 2018, nr XX/2, s. 91 i in.

przyjęto, że podstawowym kryterium doboru terminów wyjściowych artykułów hasłowych będą jednostki terminologiczne, które posiadają denotaty w pejzażu architektonicznym i urbanistycznym na terytorium Polski (lub ziem historycznie polskich). Tymi denotatami są zatem obiekty, zjawiska zaadaptowane w ramach przenikania się stylów i prądów architektonicznych (np. *konstrukcja szachulcowa*). Są to także zjawiska, będące typowymi osiągnięciami architektury polskiej (np. *wiązanie gotyckie, sklepienie piastowskie, renesans lubelski, gotyk śląski*). Korpus leksykograficzny uzupełniają jednostki leksykalne, których denotaty prymarnie odsyłają do architektury i urbanistyki innych kultur, ale wskutek wydarzeń historycznych, migracji etnicznych, kontaktów wielokulturowych przeniknęły na terytorium Polski i stały się względnie trwałym elementem architektonicznym polskiego krajobrazu (np. *molenna, cerkiew, kirkut*). Wszystkie jednostki wyjściowe otrzymują odpowiedniki rosyjskie. Co jest bardzo istotne, przyjęto tekstocentryczną formułę pozyskiwania jednostek hasłowych. Zgodnie z nią do korpusu terminów wyjściowych weszły jednostki z poświadczeniem ich użycia w konkretnych tekstach polskich (branżowych tekstach specjalistycznych: monografiach i artykułach naukowych, popularnonaukowych, tekstach użytkowych, w tym renomowanych przewodnikach turystycznych). Przedstawiona pokrótce koncepcja doboru haseł i sposób ich prezentacji uzasadniają skorzystanie w procesie ich tworzenia z elektronicznych korpusów językowych. Zgromadzone w nich zasoby tekstowe stwarzają możliwości ekscerpcji materiału słownikowego, zaś narzędzia im przypisane pomagają usprawnić ten proces.

4. ZASTOSOWANIE KORPUSÓW JĘZYKOWYCH W PROJEKCIE

Polsko-rosyjska konfiguracja językowa, objaśniająco-przekładowy charakter projektu oraz jego zakres tematyczny — te trzy parametry zdeterminowały dobór zasobów korpusowych i narzędzi, usprawniających przetwarzanie danych. Proces powstawania projektu leksykograficznego o formule objaśniająco-przekładowej przebiega w kilku etapach. Są nimi: ustalenie tzw. hasłosłownika, czyli korpusu jednostek podstawowych stanowiących bazę siatki hasłowej, stworzenie gniazdowego układu haseł, opatrzenie terminów związłą definicją, dobór optymalnych ekwiwalentów przekładowych. Ten złożony charakter czynności uzasadnia wykorzystanie w trakcie prac leksy-

kograficznych, a obecnie przeanalizowanie pod kątem użyteczności leksykograficznej zasobów różnych typów korpusów: narodowych jednojęzycznych, równoległych, zrównoważonych czy wreszcie skorzystanie z takich dostępnych narzędzi, które pomogą tworzyć korpusy na własne potrzeby i zgodnie z nimi je przetwarzać. Tak więc do korpusów językowych, których zasoby źródłowe i narzędzia ich ekstrakcji są wykorzystywane w projekcie, należą: Narodowy Korpus Języka Polskiego, Narodowy Korpus Języka Rosyjskiego, Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego, platforma Sketch Engine⁹. Poniżej zaprezentowane zostaną wykorzystane narzędzia oraz zakres ich przydatności w trakcie realizacji kolejnych etapów prac nad słownikiem.

4.1. NARODOWY KORPUS JĘZYKA POLSKIEGO (NKJP)

NKJP¹⁰ to największy referencyjny zbiór danych języka polskiego z dwoma podkorpusami¹¹. Zgodnie z przyjętym kierunkiem opracowania haseł NKJP może być pomocny na etapie tworzenia siatki haseł wyjściowych. Na potrzeby projektu skorzystano z wyszukiwarki korpusowej PELCRA, optymalnej dla przeszukiwań leksykalnych, w tym kolokacji. Umożliwia ona przeszukiwanie zasobów podkorpusu zrównoważonego stylistycznie. Zgodnie z zasadą reprezentatywności podkorpus tworzą teksty wybrane na podstawie

⁹ To nie oznacza, że tylko wymienione narzędzia zostały uwzględnione na wstępnym etapie doboru zasobów elektronicznych. Przetestowano także inne korpusy, m.in. korpus równoległy InterCorp, będący częścią Narodowego Korpusu Języka Czeskiego (www.intercorp.korpus.cz), korpus równoległy ParaSol, zawierający dane głównie języków słowiańskich (www.slavist.de). W obu przypadkach jednak kolekcje tekstów specjalistycznych dla polsko-rosyjskiej konfiguracji językowej są bardzo słabo reprezentowane lub w ogóle nie są notowane. Także wyszukiwarka korpusowa KonText w sieci Clarin-PL z dostępnym w niej polsko-rosyjskim korpusem równoległym nie przyniosła oczekiwanych rezultatów z powodu bardzo niskiej reprezentatywności tekstów branżowych (<https://kontext.clarin-pl.eu>).

¹⁰ <http://www.nkjp.uni.lodz.pl>

¹¹ Struktura NKJP, zasady użytkowania i możliwości zastosowania zostały omówione w pracach: A. Przepiórkowski, M. Bańko, R.L. Górski, B. Lewandowska-Tomaszczyk, *Narodowy Korpus Języka Polskiego*, Wydawnictwo PWN, Warszawa 2011; także: R. Bańko, M.R. Górski, *Praktyczny przewodnik po korpusie języka polskiego*, w: M. Hebal-Jezińska (red.), *Praktyczny przewodnik po korpusach języków słowiańskich*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2014.

ustalonych proporcji rejestrów stylistycznych oraz rodzajowych. W NKJP przyjęto następujące proporcje w oparciu o badania czytelnictwa w Polsce: książki – 29%, prasa – 50%, Internet – 7%, teksty mówione – 10%, inne teksty – 4%. Taka typologia, zupełnie słuszna z punktu widzenia ogólnego przeznaczenia NKJP, nie zapewnia jednak optymalnej reprezentacji tekstów specjalistycznych jako potencjalnego źródła terminów z zakresu architektury i urbanistyki. Skorzystano zatem z dostępnej w zasobach NKJP kolekcji tekstów użytkowych (informatatorów, przewodników turystycznych), prasowych i popularnonaukowych. Funkcjonalność wyszukiwarki PELCRA na potrzeby niniejszego projektu zostanie zaprezentowana na przykładzie terminu *elewacja* – jednego z bazowych pojęć z zakresu architektury.

Tabela 1. Konkordancje z udziałem leksemu *elewacja*
(źródło: www.nkjp.uni.lodz.pl)

Fragment tekstu	Wystąpienie leksemu	Źródło
1. ...płachina z XIV w. Kościół (dammyj szpitalny) z I poł. XVII w.	elewacja	Wierusz. Mazury, Suwa...
2. ...	elewacja	Suy - kłhez
3. ...	elewacja	Parada radców
4. ...	elewacja	Parada radców
5. ...	elewacja	Znane i nieznane: r...
6. ...	elewacja	Znane i nieznane: r...
7. ...	elewacja	Znane i nieznane: r...
8. ...	Elewacja	Znane i nieznane: r...
9. ...	Elewacja	Znane i nieznane: r...
10. ...	elewacja	Znane i nieznane: r...
11. ...	elewacja	Znane i nieznane: r...
12. ...	elewacja	Znane i nieznane: r...
13. ...	elewacja	Znane i nieznane: r...
14. ...	elewacja	Znane i nieznane: r...
15. ...	elewacja	Znane i nieznane: r...
16. ...	Elewacja	Znane i nieznane: r...
17. ...	elewacja	Znane i nieznane: r...
18. ...	elewacja	Znane i nieznane: r...
19. ...	elewacja	Świat dla opozycj...
20. ...	elewacja	Góry Śląsk
21. ...	elewacja	Góry Śląsk
22. ...	elewacja	Góry Śląsk
23. ...	elewacja	Góry Śląsk
24. ...	elewacja	Góry Śląsk
25. ...	elewacja	Góry Śląsk
26. ...	elewacja	Góry Śląsk
27. ...	Elewacja	Góry Śląsk
28. ...	elewacja	Przewodniki po Polsc...
29. ...	elewacja	Przewodniki po Polsc...
30. ...	elewacja	Przewodniki po Polsc...
31. ...	Elewacja	Przewodniki po Polsc...
32. ...	Elewacja	Przewodniki po Polsc...
33. ...	Elewacja	Przewodniki po Polsc...

Powyższy zrzut ekranu ilustruje przykład konkordancji w najpopularniejszej formie KWIC, a więc wyszukiwany wyraz, wyróżniony na środku generowanej strony, jest przedstawiony w otaczającym go kontekście z lewej i prawej strony¹². Wynik przeszukiwań to łącznie 382 akapity. Dzięki anotacjom zewnętrznym możliwe jest pozyskanie informacji o typach tekstów z udziałem interesującego nas leksemu *elewacja*, a zatem i stopniu ich różnorodności, co ilustruje poniższa tabela:

¹² B. Lewandowska-Tomaszczyk, *Podstawy...*, s. 297.

Tabela 2. Profil występowania dla zapytania: *elewacja*
(źródło: www.nkjp.uni.lodz.pl)

#	Typ	Liczba wystąpień (A)	Słowa w kategorii (B)	A/(B/1M)
1	publ	326	129,475,805	2.518
2	inf_por	18	8,078,242	2.228
3	fakt	15	14,874,971	1.008
4	net_interakt	11	9,218,718	1.193
5	nd	4	8,680,759	0.461
6	konwers	3	1,774,144	1.691
7	lit	3	30,230,684	0.099
8	qmow	2	23,323,854	0.086
9	lit_dramat	0	48,374	0
10	lit_poezja	0	74,793	0
11	lit_proza	0	4,512,240	0

Przedstawione liczby wydają się miarodajnym sygnałem użyteczności tych danych do celów leksykograficznych. Jaką zatem wartość ma ta informacja? Po pierwsze wskazuje ona na potencjał kolokacyjny wyjściowego terminu *elewacja*. Informacje pozyskane z modułu ekstrakcji kolokacji są istotne dla gniazdowego układu hasel, pozwalają bowiem znacznie rozszerzyć zbiór podhasel jako terminów złożonych derywowanych jednoskładnikowym bazowym terminem *elewacja*. Z powyższego zbioru konkordancji, jak również wspomagającej proces pozyskiwania kolokacji wyszukiwarki *Kolikator* udało się wyszukać następujące połączenia wyrazowe z bazowym leksemem *elewacja*: *frontowa elewacja* (4 wystąpienia w korpusie zrównoważonym (KZ), 19 w korpusie ogólnym (KO)), *elewacja frontowa* (14 użyć w KZ, 84 – w KO), *elewacja czołowa* (2 użycia w KZ, 3 – w KO), *elewacja ogrodowa* (4 użycia w KZ, 11 – w KO), *ogrodowa elewacja* (0 użyć w KZ, 1 – w KO), *zachodnia elewacja* (2 użyć w KZ, 8 – w KO), *elewacja zachodnia* (1 użycie w KZ, 27 – w KO) *barokowa elewacja* (2 użyć w KZ, 3 – w KO). Te liczby, w zależności od ich wartości, wskazują na stabilność połączeń wyrazowych (i zarazem poświadczają istnienie terminu złożonego, np. *elewacja frontowa/frontowa elewacja*) bądź jedynie sygnalizują potencjał kolokacyjny wyrażenia (*elewacja czołowa*), wymagający dodatkowych poświadczeń w innych źródłach. Powyższe dane dostarczają także ważnych informacji na temat wariantywności szyku komponentów połączeń kolokacyjnych (typu *frontowa elewacja/elewacja frontowa*, *ogrodowa elewacja/elewacja ogrodowa*). Większość hasel podrzędnych,

tj. terminów złożonych, ma strukturę imienną (połączenie rzeczownika z zależnym od niego semantycznie przymiotnikiem jako kolokatem). Z reguły, zgodnie z normą języka polskiego forma przymiotnika występuje w postpozycji w stosunku do rzeczownika, ale nie zawsze: niekiedy pojawia się szyk odmienny. Zbadanie frekwencji takich połączeń imiennych, po pierwsze, pozwala zasygnalizować ten problem, po drugie, pomaga w ustaleniu, który z wariantów szyku wykazuje tendencje dominujące, który natomiast spotykany jest rzadziej. Może także na podstawie frekwencji wystąpień sugerować równorzędność obu wariantów. Obecność anotacji zewnętrznych pozwala przeanalizować, w jakim stopniu dany termin jest stosowany niezależnie przez różnych autorów. Im większa różnorodność tekstów, tym silniejszy potencjał kolokacyjny i zarazem terminologiczny badanego wyrażenia.

Możliwość przejścia w przeglądarce NKJP od widoku konkordancji do widoku rozwinięcia otoczenia tekstowego badanej jednostki w praktyce leksykograficznej oznacza pozyskanie dodatkowego, stosunkowo łatwego w obsłudze źródła potencjalnych nowych hasel. Potwierdza to następujący przykład:

Tabela 3. Przykład występowania kontekstowego terminu złożonego *elewacja ogrodowa* (źródło: www.nkjp.uni.lodz.pl)

Dane tekstu dla wyniku:	8
Identyfikator skłonu:	IPPAN_p0000106137
Tytuł:	Czasz. <i>elewacja ogrodowa</i>
Wydawca:	Przedsiębiorstwo Wydawnicze Produkcyjne Ushagow Sp. z o.o.
Źródło:	Typodruk Regionalny "Gazeta Ciepłochowska" nr 16 (80)
Kanał:	Hasła i gram. typodruk
Typ:	Hyp_p001
Autorzy:	UG,
Data publikacji:	2007-04-19

Dziękuję! Katarzyna Suchanowicz
 *Budynek anegdotycki o elementach neobarokowych, pocz. XX w. (przed 1914). Dwupiętrowy z nielicznymi interesami, piętra dwuspójniasto-, parter trzynastosienny, z sienią przziadkową na stronie osi, kryta przeszklona sklepieniem odcinkowym z żelaznym zebrowaniem o składowiu krzyżowym, z gurtami spływającymi na przyciemnione przeszklenie, z zachowanymi resztkami emali. W klatce schodowej drzwi do mieszkan i podłazki białoczerwone, z neobarokowymi uszporami. Balkony z schodów neobarokowe, szklana o smolech cegrowo-skałkowych, od frontu analogiczne balustrady, przy schodach do interes. **Elewacja frontowa** z poziomymi ryzalitami skrajnymi oraz dwukondygnacyjnym wykuszem na osi środkowej, ponad klatką trójkatnie szczyty z marmurami - białymi, z dekoracją neogotycką w formie listewki (obierających między innymi skłani, sterczy i Fryzów arkadowych, zwieńczona kolumnami muretkowymi balustrady z kolumnami balustradami balustradami w ryzalitu skrajnym, nad oknami drugiego piętra także w otoczeniu neogotyckim składowym. Brama wjazdowa o kaskadownych węgarkach z żelaznymi ochronnymi rozetami i półokrągłymi odbojami, w osi drzwi żaluzje dwukondygnowe z zakratowanymi przeszkloniami o motywach geometrycznych i muretkowych rzeźbami. Ciepłota zachowana pierwotna stolarka. Dach mansardowy z wyskokami.
 Katalog zabudów sztuki. Miasto Ciepłochowa, pod red. Zofii Roznow i Ewy Smalukowskiej, cz. 1, Instytut Sztuki Polskiej Akademii Nauk, Warszawa 1995.

Nawet pobieżna analiza poświadczenia tekstowego hasła *elewacja frontowa* (wyróżnionego graficznie kolorem żółtym) pomaga wyszukać w ciągu zespolonych wyrażen nowe artykuły hasłowe: *sklepienie* wraz z podhasłem *odcinkowe sklepienie zebrowe*, *gurt*, *pilastry*, *sztukaterie*, *laskowanie*, *ryzalit*, *fryz arkadowy*, *dach mansardowy* itp. Wprawdzie w NKJP nie ma możliwości hipertekstowego przełączenia się na te jednostki, to jednak, mimo braku tej funkcji, pojawia się stosunkowo łatwa pomoc w odnotowaniu i ewentualnym posze-

rzeniu korpusu haseł wyjściowych: bazowych i od nich pochodnych, zgodnie z zadeklarowaną orientacją tekstocentryczną prac nad słownikiem. Zatem korpus jednojęzyczny NKJP można wykorzystać jak źródło poszukiwań bazowych haseł wyjściowych oraz pochodnych od nich, służących rozszerzeniu artykułu hasłowego. Można potraktować go także jako źródło informacji o strukturze komponentów kolokacji imiennych.

4.2. NARODOWY KORPUS JĘZYKA ROSYJSKIEGO (NKJR)¹³

Jest to reprezentatywny i zbalansowany zbiór tekstów literatury pięknej, tekstów prasowych, naukowych i popularnonaukowych, wypowiedzi mówionych, tekstów multimedialnych, korespondencji, dzienników itd¹⁴. Na potrzeby eksploracji materiału słownikowego wykorzystano korpus ogólny (podstawowy) oraz prasowy. Fakt, że korpus ten odnosi się do zasobów języka rosyjskiego (skądinąd bardzo bogatych i różnorodnych pod względem typów tekstów w nim zawartych), wskazuje na jego możliwe zastosowanie w pozyskiwaniu informacji o ekwiwalentach rosyjskich, a więc będzie pomocny w uzupełnieniu prawej strony siatki hasłowej. Tu pojawia się jednak pytanie o procedurę takiej ekstrakcji. Brak w korpusie danych wyjściowych, tj. haseł w języku polskim znacznie utrudnia proces pozyskiwania potrzebnych informacji, ale go nie uniemożliwia. Funkcja konkordancji w formacie KWIC, naturalnie dostępna także w NKJR, daje możliwość łatwego wglądu w rzeczywiste użycia terminów w ich najbliższych kontekstach. A te mogą ujawniać pewne tendencje połączeń wyrazowych. Ponadto, dysponując już pewnym zestawem złożonych terminów imiennych (tj. połączeń rzeczownik+przymiotnik), można skorzystać z posiadanej wiedzy i kompetencji językowych i dokonać prób własnego, dosłownego tłumaczenia takich kolokacji, a następnie poszukać potwierdzenia poprawności ich przekładu w rosyjskich zasobach korpusowych. Prześledźmy tę procedurę na przykładzie

¹³ www.ruscorpورا.ru

¹⁴ Szczegółowe informacje nt. NKJR m.in. w pracach: Д.О.Добровольский, А.А.Кретов, С.А. Шаров, *Корпус параллельных текстов: архитектура и возможности использования*, w: *Национальный корпус русского языка 2003–2005*, Индрик, Москва 2005; M. Duszkin, *Praktyczny przewodnik po korpusie języka rosyjskiego*, w: M. Hebał-Jezińska (red.), *Praktyczny przewodnik po korpusach języków słowiańskich*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2014.

leksemu *фасад* jako wcześniej ustalonego (na podstawie innych źródeł leksykograficznych) odpowiednika pol. terminu *elewacja*.

Tabela 4. Zestawienie terminów złożonych z członem nadrzędnym *elewacja* i propozycje ich rosyjskich odpowiedników (opracowano na bazie danych dwutekstów i NKJR)

Termin w języku polskim	Termin w języku rosyjskim	Korpus ogólny: liczba poświadczeń	Korpus prasowy: liczba poświadczeń
elewacja boczna	боковой фасад (A)	8	5
elewacja frontowa	парадный фасад (D)	8	0
elewacja podwórkowa	дворовый фасад (D)	0	1
elewacja szklana	стеклянный фасад (A)	3	6
elewacja ślepa	слепой фасад (?) (A)	0	0
elewacja tylna	задний фасад (D)	17	5
elewacja wentylowana	вентилируемый фасад (D)	0	3
elewacja zachodnia	западный фасад (A)	7	1

Propozycje terminów rosyjskojęzycznych — to rezultat tłumaczenia własnego (symbol A) lub wynik przeszukiwań dwutekstów (polskich przewodników turystycznych i ich rosyjskich przekładów (symbol D), w których z reguły udało się odnaleźć jednostkowe użycia, wymagające kolejnych poświadczeń. Niektóre z nich pozyskano z zasobów NKJR (*боковой фасад*, *парадный фасад*, *стеклянный фасад*, *западный фасад*), inne wymagają kolejnych poszukiwań (*дворовый фасад*, *слепой фасад*). Ich funkcjonowanie w kontekście, nie zaś w formie wyizolowanej pomaga także ocenić, w jakim stopniu są to jednostki ekwiwalentne pod względem semantycznym i funkcjonalnym (np. pol. *elewacja frontowa* i ros. *парадный фасад*).

Warto nadmienić, że w ramach podkorpusek w NKJR istnieje polsko-rosyjski korpus równoległy, zawierający polskie teksty oryginalne (w znakomitej większości literackie — utwory Andrzeja Sapkowskiego, Bolesława Prusa, Stanisława Lema, Stefana Żeromskiego i in.). Ich przydatność jest wprawdzie stosunkowo niewielka dla konkretnych działań leksykografii przekładowej, jednak odnotowano nieliczne przykłady propozycji translatorskich, które mogą dopełnić podstawowe ekwiwalenty systemowe. W charakterze przykładu posłużę się dwoma cytatami z anotowanej w podkorpuse powieści Andrzeja Sapkowskiego *Narrentum* i jej przekładu:

Tabela 5. Przykłady użycia leksemu *empora* w powieści *Narrentum* i sposoby jego tłumaczenia na język rosyjski (źródło: www.ruscorpora.ru)

Wersja polska	Wersja rosyjska
Reszta prezbiterium, cała nawa, drewniane empory i stalle tonęły w mętym półmroku.	Остальная часть пресвитерии, весь неф, деревянные эмпоры и сталлы тонули в туманной полутьме.
— W kościele sam luksus — Drosselbart wyłamał z empory w prezbiterium kolejną deskę.	— В церкви сплошная роскошь. — Дроссельбарт выломал из галереи в пресвитерне очередную доску.

Powyzsze przykłady propozycji translatów, choć są autorstwa tego samego tłumacza, to jednak sygnalizują funkcjonowanie w języku rosyjskim dwóch ekwiwalentów polskiego leksemu *empora*. Zresztą poprawność użycia leksemu znajduje potwierdzenie w fakcie występowania w języku polskim synonimu wyrazu *empora* — jest nim właśnie *galeria*. Chociaż zasoby dwutekstów (w kontekście potrzeb leksykograficznych) w NKJR są skromne w stosunku do realnych potrzeb leksykograficznych, to jednak warto je wykorzystać łącznie z zasobami innego korpusu, tego samego typu, ale tworzonego na gruncie polskim. Jest nim Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego.

4.3. KORPUS POLSKO-ROSYJSKI UNIwersYTETU WARSZAWSKIEGO¹⁵ (KPRUW)

To korpus reprezentatywny, równoległy anotowany morfosyntaktycznie. Jego bazę tworzą w głównej mierze teksty literackie (90%), 5% — teksty prasowe, 4% — teksty religijne¹⁶. Odnotowano w niej znaczny udział powieści historycznych, co stwarza perspektywy prześledzenia terminów z zakresu architektury zamkowej, obronnej, także kościelnej. W niniejszym opracowaniu, stosownie do poczynionych wcześniej ustaleń, dokonam niewielkiej próbki porównania propozycji translatorskich zamieszczonych w obu korpusach w celu wskazania — poprzez takie zestawienie — na ich wspólną większą użyteczność.

¹⁵ www.pol-ros.polon.uw.edu.pl

¹⁶ Więcej szczegółów na temat Korpusu w pracy: M. Łaziński, M. Kuratczyk, *Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego*, w: E. Gruszczyńska, A. Leńko-Szymańska (red.), *Polskojęzyczne korpusy równoległe / Polish-language Parallel Corpora*, Instytut Lingwistyki Stosowanej WLS, Warszawa 2016, s. 85–87.

Tabela 6. Przykłady przekładu terminu *prezbiterium* na język rosyjski w kolekcjach tekstów w NKJR oraz KPRUW(źródło: www.ruscorpora.ru; <http://pol-ros.polon.uw.edu.pl>)

Wersja oryginalna polska	Wersja rosyjska
Reszta prezbiterium , cała nawa, drewniane empory i stale tonęły w mętным półmroku [Narrenturm, Sapkowski Andrzej] (NKJR).	Остальная часть пресвитерии , весь неф, деревянные эмпоры и стали тонули в туманной полутьме.
W stronie prezbiterium utworzyła się zbita masa «przystępujących» do antałka z winem [Popioły, Żeromski Stefan] (KPRUW).	У алтаря образовалась плотная толпа солдат, «со страхом и верою приступавших» к бочонку с вином [Пепел, Жеромский Стефан].
Nie pozostało nic innego, tylko wrócić do prezbiterium i szukać tam pożywienia. [Popioły, Żeromski Stefan] (KPRUW).	Цедро не оставалось ничего другого, как вернуться в алтарь и поискать там какой-нибудь снеди [Пепел, Жеромский Стефан].
Weszli do ciemnej kruchty i Witek zobaczył przy końcu nawy, pod balaskami prezbiterium , dwa rzędy płonących świec [Kronika wypadków miłosnych, Konwicky Tadeusz] (KPRUW).	Вошли в темный притвор, и Витек увидел в глубине нефа, под сиянием пресвитерия , два ряда горящих свечек [Хроника любовных происшествий, Конвицкий Тадеуш].

Wyróżniony w tabeli leksem *prezbiterium* ma kilka zaproponowanych przez różnych tłumaczy odpowiedników przekładowych. Są nimi: *пресвитерия*, *алтарь* i *пресбитерий*. Takie wyniki naprowadzają leksykografa, po pierwsze, na możliwą wariantowość translatu *пресвитерия/пресбитерий*, po drugie, sugerują możliwość zastosowania także leksemu *алтарь* jako funkcjonalnego odpowiednika terminu *prezbiterium*:

4.4. PLATFORMA KORPUSOWA SKETCH ENGINE

Sketch Engine¹⁷ jest komercyjnym narzędziem korpusowym, coraz szerzej stosowanym w światowej leksykografii¹⁸. Zapewnia dostęp do ponad 500 zbiorów tekstów w przeszło 90 językach¹⁹. Oprócz boga-

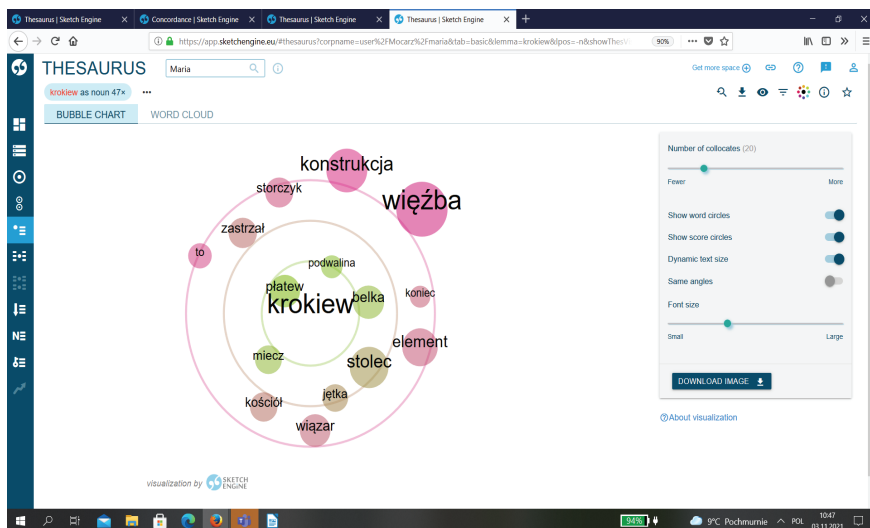
¹⁷ www.sketchengine.eu

¹⁸ Szerzej na temat platformy w pracy A. Kilgarriff, *The Sketch...*, s. 7–36, także: E.B. Работалова, *Корпусные технологии в лексикографии (на примере английской терминологии нефтепереработки)*, „Вестник Московского государственного областного университета, Серия: Лингвистика” 2020, nr 5, s. 8–18.

¹⁹ Na potrzeby projektu, jak również niniejszego opracowania skorzystano z bezpłatnego dostępu do platformy Sketch Engine w ramach licencji uzyskanej przez

projekcie jednak rzeczywista przydatność tej funkcji na bazie ogólnego korpusu Polish Web 2109 okazała się nieco mniejsza z uwagi na uzyskanie dostępu do platformy na nieco późniejszym etapie prac nad słownikiem, kiedy znakomita większość z zamieszczonych w tzw. chmurze leksemów została już wprowadzona do struktury słownika. Niemniej jednak z funkcji tezauryś do ekstrakcji terminów z własnego subkorpusu utworzonego na bazie tekstów naukowych z zasobami terminologii o wyższym stopniu specjalizacji niż obecna w tekstach tworzących korpus Polish Web 2019 przyniosło satysfakcjonujące rezultaty. Zostaną one zaprezentowane na przykładzie terminu *krokiew*:

Tabela 8. Tezaurusz z hasłem wyjściowym *krokiew*, stworzony na bazie własnego subkorpusu (źródło: www.sketchengine.eu)



Tym razem uzyskane wyniki z wykorzystaniem funkcji tezauryś znalazły zastosowanie w procesie konstruowania siatki hasłowej. Na ich podstawie do korpusu hasel włączono następujące bazowe terminy (oznaczające elementy konstrukcji dachowych): *zastrzał* (33 użycie w subkorpusie), *stołec* (76), *jętka* (25), *wiązar* (41), *miecz* (28), *storczyk* (32), *więźba* (154). Następnie część z tych terminów bazowych stała się podstawą układu gniazdowego, grupującego terminy złożone. Tu, analogicznie do działań w NKJP, odwołano się do konkordancji oraz tzw. funkcji Sketch Word — jednej z podstawowych funkcji platformy dającej wgląd w możliwości dystrybucyjne wyszukiwanej

frazy. Poniżej zaprezentowana zostanie pierwsza strona zarejestrowanych połączeń leksykalnych dla terminu *więźba*:

Tabela 9. Możliwości dystrybucyjne terminu *więźba* na bazie własnego subkorpusu (źródło: www.sketchengine.eu)

a_modifier	prec_prep	post_verb	prec_verb	post_o	is_obj
jętkowy Węźby jętkowe	w w więźbach	spotykać więźby spotykamy	nazywać nazywamy więźbą	stolec węźby o stołcach	wykazać węźby ostate wykazać
otwarty otwarta więźba	przy Przy więźbach	mieć węźbach żytkowych mamy	tworzyć tworzą więźby płatwiwo-kleszczowe	redukja węźbach o redukcji	opuszcili węźby brukowej opuszczili
storczykowy węźba storczykową zredukowaną	od od więźb	nazywać nazywają będziemy nazywali	wykonstruować wykonstruowali więźbę	przekrój węźby o masyjnych przekrojach	pozwałać węźby wolnej pozwala
taki W takich więźbach	obok Obok powyższych więźb	upodobić węźby upodobiłają	pokazywać pokazują więźbę		nazywać węźbę będziemy nazywali
płatwiwo-kleszczowych podgrano węźb płatwiwo-kleszczowych	jak jak więźba	wykazać węźby ostate wykazać	składać składają się więźba		
płatwiwo-kleszczowe Wszystkie płatwiwo-kleszczowe	do do wstępnych więźb	opuszcili węźby brukowej opuszczili	istnieć istnieć także więźba		
dachowy Węźby dachowe	nad nad więźbami	polegać węźbach wieszarowych polega	nazwać nazwać więźbę		
półotwarty węźby półotwarte	o o więźbach	okazać okazują okazy			
płatwiwo-kleszczowa Węźba płatwiwo-kleszczowa o stołcach		stanowiąc wstępnym płatwiwo-kleszczowych stanowiąc			
wolny Węźby wolne		pozwałać węźby wolnej pozwala			
wieszarowy węźbach wieszarowych		składać składają się			
płatwiwo-kleszczowa węźba płatwiwo-kleszczowa		tworzyć węźby płatwiwo-kleszczowych tworzą			
jednostorczykowa węźba jednostorczykowa					

Narzędzie to pozwoliło na wyszukanie i umieszczenie w strukturze siatki hasłowej jako podhasła następujących terminów złożonych o strukturze imiennej: *więźba dachowa jętkowa*, *więźba storczykowa*, *więźba stolcowa*, *więźba płatwiwo-kleszczowa*, *więźba otwarta*, *więźba wieszarowa*, *więźba wolna*.

Korpus w języku rosyjskim, dostępny na platformie Sketch Engine, skłania do pytania o jego przydatność w poszukiwaniu odpowiedników po prawej stronie artykułu hasłowego. Korpus ten jest niezależny, cechy korpusu zrównoważonego w stosunku do wersji polskiej wykazuje w stopniu minimalnym, jeśli przyjąć za kryterium ich porównania podobieństwo typów tekstów takie korpusy tworzących. Kolekcje tekstów rosyjskojęzycznych nie reprezentują gatunków wysoko specjalistycznych, zatem trudno w nich szukać ekwiwalentów takich polskich terminów, jak np. *więźba jętkowa*. Niemniej podejmowane próby poszukiwania ekwiwalentów mają wartość sondażową, podobną do tej, jaka została zaprezentowana wcześniej przy analizie przydatności zasobów NKJR. Ponownie wykorzystany autorski zbiór kolokacji z komponentem bazowym *φαcαδ* dostarczył tym razem następujących wyników (Tab. 11).

Tabela 10. Zestawienie terminów złożonych z członem nadrzędnym *elewacja* i ich rosyjskich odpowiedników (opracowano na bazie danych Sketch Engine)

Termin w języku polskim	Termin w języku rosyjskim	Korpus rusTenTen11 liczba poświadczeń
elewacja boczna	боковой фасад (A)	1722
elewacja frontowa	парадный фасад (D)	961
elewacja podwórkowa	дворовый фасад (D)	5
elewacja szklana	стеклянный фасад (A)	2687
elewacja ślepa	слепой фасад (A)	17
elewacja tylna	задний фасад (D)	981
elewacja wentylowana	вентилируемый фасад (D)	10
elewacja zachodnia	западный фасад (A)	3538

Powyższe dane potwierdzają stabilność połączeń wyrazowych, a tym samym status terminów oraz trafność doboru rosyjskich ekwiwalentów. Pozyskane dane, w szczególności jeśli pochodzą z więcej niż jednego korpusu i są zespolone ze sobą, stają się weryfikatorem poprawności odpowiednika przekładowego: zaproponowanego wcześniej apriorycznie (tj. na podstawie posiadanej kompetencji lingwistycznej) lub odnotowanego w innym niż zasoby korpusowe źródle (dwutekście).

5. WNIOSKI

Implementacja zasobów korpusowych w polsko-rosyjskim projekcie leksykograficznym *Architektura i urbanistyka* przebiega z różną intensywnością na poszczególnych etapach jego opracowywania. Najbardziej pomocne są korpusy językowe (wraz z dostępnymi narzędziami) na pierwszych etapach, tj. budowania hasłownika (hasel podstawowych i pochodnych od nich). W znacznie mniejszym stopniu zaprezentowane korpusy językowe są wykorzystywane w procesie poszukiwania odpowiedników rosyjskich. Dzieje się tak za sprawą stosunkowo ubogiej kolekcji tekstów specjalistycznych, a nawet użytkowych dostępnych w korpusach równoległych (PRKUW oraz PRKR). Często uzyskane wyniki mają wartość jedynie sondażową, pomagają zwrócić uwagę na leksem/kolokację jako potencjalny materiał leksykograficzny. Jednak dla uzyskania pewności, że taka jednostka nosi znamiona terminu, często wymagana jest większa liczba poświadczeń tekstowych. Istotne jest łączenie danych pozyskanych

z różnych zasobów korpusowych w celu otrzymania informacji relevantnych dla zawartości artykułu hasłowego projektowanego słownika. Jak wykazano, łączenie dostępnych zasobów zwiększa ich efektywność w trakcie prac leksykograficznych zarówno na etapie doboru haseł wyjściowych (w szczególności terminów złożonych), jak i na etapie poszukiwań ekwiwalentów. Zasadniczo w kontekście konkretnych potrzeb leksykograficznych można mówić o ich komplementarnym charakterze. Sposobem rokującym zwiększenie zasobności korpusów jest tworzenie własnych subkorpusów z tematycznie i gatunkowo sprofilowanymi kolekcjami tekstów dostosowanymi do indywidualnych potrzeb badawczych. Jednocześnie przedstawiony zakres możliwości wykorzystania dostępnych korpusów językowych i ich narzędzi potwierdza zasadność przyjętej metody analizy materiałów wspomaganej korpusami (*corpus-assisted analysis*), w której dane korpusowe jedynie wspomagają prace leksykograficzne. Zaprezentowane studium przypadku, oparte na badaniach empirycznych, z całą pewnością wskazuje na nową jakość w pracy leksykografa przy wykorzystaniu danych i narzędzi korpusowych.

REFERENCES

- Baker, Paul. *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press, 2010.
- Bańko, Mirosław, Górski Rafał. "Praktyczny przewodnik po korpusie języka polskiego." Hebal-Jeziarska Milena. Ed. *Praktyczny przewodnik po korpusach języków słowiańskich*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2014: 11–26.
- Bralewski, Dariusz. *Od przekładu do słownika. Korpus równoległy w redakcji słowników tłumaczeniowych*. Łask: Oficyna Wydawnicza LEKSEM, 2012.
- Charciarek, Andrzej. "Korpus równoległy InterCorp w leksykografii przekładowej – możliwości i ograniczenia." *Roczniki Humanistyczne* 2019, t. LXVII, 7: 79–92.
- Dobrovol'skiy, Dmitriy Olegovich. "Korpusy tekstov i dvuyazychnaya frazeografiya." *Vestnik Novosibirskogo gosudarstvennogo pedagogicheskogo universiteta*, 2015, 5(27): 23–37. [Добровольский Дмитрий Олегович. "Корпусы текстов и двуязычная фразеография." *Вестник Новосибирского государственного педагогического университета*, 2015, 5(27): 23–37]. DOI: 10.15293/2226–3365.1505.02.
- Dobrovol'skiy, Dmitriy Olegovich, and Kretov, Aleksey Aleksandrovich, and Sharov, Serney Aleksandrovich. "Korpus parallel'nykh tekstov: arkhitektura i vozmozhnosti ispol'zovaniya." *Natsional'nyy korpus russkogo yazyka: 2003–2005*. Moskva: Indrik, 2005: 263–296 [Дмитрий О. Добровольский, and Кретов Алексей Александрович, and Шаров Серней Александрович. "Корпус параллельных текстов: архитектура и возможности

- использования.” *Национальный корпус русского языка: 2003–2005*. Москва: Индрик, 2005: 263–296; <https://ruscorpora.ru/new/corpora-biblio.html> (12.08.2021)].
- Duszkin, Maksim. “Praktyczny przewodnik po korpusie języka rosyjskiego.” Hebal-Jeziarska Milena. Ed. *Praktyczny przewodnik po korpusach języków słowiańskich*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2014: 142–165.
- Faaß, Gertrud. “Lexicography and Corpus Linguistics.” Fuertes-Olivera, Pedro. A. Ed. *The Routledge Handbook of Lexicography*. Oxon: Routledge, 2017: 123–137.
- Grabowski, Łukasz. “Statystyka, programowanie i co jeszcze? O wyzwaniach współczesnego językoznawstwa.” Mokijenko, Walerij, and Tarsa, Jadwiga. Eds.. *Język i pamięć. Księga jubileuszowa dedykowana Pani Profesorowi Wojciechowi Chlebdzie z okazji 70. urodzin*. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2020: 257–271.
- Grabowski, Łukasz. “Korpusy dwu- i wielojęzyczne w służbie tłumacza, leksykografa i badacza: poszukiwanie ekwiwalentów przekładowych w świetle hipotez dotyczących istnienia uniwersaliów tłumaczeniowych.” Chlebda, Wojciech. Ed. *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2011: 89–112.
- Halliday, Michael. A.K., and Teubert, Wolfgang, and Yallop, Colin, and Cermáková Anna. *Lexicology and Corpus Linguistics. An Introduction*. Londyn: Continuum, 2004.
- Hanks, Patrick. “The corpus revolution in lexicography.” *International Journal of Lexicography*, 2012, no. 25(4): 398–436.
- Kiklewicz, Aleksander. “Korpus internetowy jako źródło informacji lingwistycznej: ograniczenia.” *Prace Językoznawcze* 2018, t. XX (2): 75–97.
- Kilgarriff, Adam, and Baisa, Vít and Bušta, Jan, and Jakubiček, Miloš, and Kovář, Vojtěch, and Michelfeit, Jan, and Rychlý, Pavel, and Suchomel, Vít. “The Sketch Engine: Ten Years On”, *Lexicography* 2014, no. 1: 7–36. DOI:10.1007/s40607-014-0009-9.
- Lewandowska-Tomaszczyk, Barbara. *Podstawy językoznawstwa korpusowego*. Łódź: Wyd. Uniwersytetu Łódzkiego, 2005.
- Łaziński, Marek, and Kuratczyk, Magdalena. “Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego.” Gruszczyńska, Ewa, and Leńko-Szymańska, Agnieszka. Eds. *Polskojęzyczne korpusy równoległe / Polish-language Parallel Corpora*. Warszawa Instytut Lingwistyki Stosowanej WLS, 2016: 83–95.
- Przepiórkowski, Adam, and Bańko, Mirosław, and Górski, Rafał L., and Lewandowska-Tomaszczyk, Barbara. *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo PWN, 2012.
- Rabotalova, Yekaterina Vladimirovna. “Korpusnyye tekhnologii v leksikografii (na primere angliyskoy terminologii neftepererabotki).” *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta, Seriya: Lingvistika*, 2020, no. 5: 8–18 [Работалова, Екатерина Владимировна. “Корпусные технологии в лексикографии (на примере английской терминологии нефтепереработки). *Вестник Московского государственного областного университета, Серия: Лингвистика*, 2020, no. 5: 8–18].
- Vetulani, Grażyna. “Problemy i korzyści wynikające z automatycznego przetwarzania korpusów — na przykładzie badań z zakresu predykcji rzeczownikowej w języku polskim.” *Roczniki Humanistyczne* 2013, t. LXI, z. 8: 13–24.

Żmigrodzki, Piotr (i in.). *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*. Kraków: Instytut Języka Polskiego PAN, 2018: 19–20.

ONLINE SOURCES

www.ruscorpora.ru (24.08.2021)
www.pol-ros.polon.uw.edu.pl (24.08.2021)
www.nkjp.pl (02.09.2021)
www.nkjp.uni.lodz.pl (04.09.2021)
<http://dict.ruslang.ru/magn.php> (02.08.2021)
www.intercorp.korpus.cz (12.09.2021)
www.slavist.de (12.09.2021)
<https://kontext.clarin-pl.eu> (20.10.2021)
www.sketchengine.eu (05.09.2021)



ROMAN ROSZKO

 ORCID: <https://orcid.org/0000-0002-2291-6939>

Instytut Sławistyki Polskiej Akademii Nauk

RUSYCYSTYCZNE ZASOBY I NARZĘDZIA CLARIN-PL

RUSSIAN LANGUAGE RESOURCES AND TOOLS ON THE CLARIN-PL WEBSITE

In this article I present multilingual resources with Russian language already created and currently being constructed by CLARIN-PL consortium. I also introduce the tools and services offered by this consortium for users interested in analysing Russian-language resources.

1. WSTĘP

Zespoły badawcze Instytutu Sławistyki Polskiej Akademii Nauk od lat 70. minionego wieku prowadzą badania kontrastywne języków słowiańskich i bałtyckich. Początkowo badania materiałowe były poprzedzane żmudną ręczną ekscerpcją zasobów drukowanych z wybranych do analizy języków. Wiązało się to z brakiem jakichkolwiek wielojęzycznych kolekcji cyfrowych, tym bardziej narzędzi do ich przetwarzania. Tę niedogodność próbowano na różne sposoby pokonać. Już w XX wieku powstał pierwszy *Eksperymentalny polsko-litewski korpus tekstów dwudziestowiecznych*¹ autorstwa Danuty Roszko i Romana Roszki, który znalazł zastosowanie w wielu studiach nad okre-

¹ Szczegółowy opis projektu jest zawarty w D. Roszko, R. Roszko, *Polsko-litewskie korpusy IS PAN i CLARIN-PL*, w: N. Birgiel, D. Roszko (red.), *Prace Bałtyckie – 7. Język. Kultura. Literatura*, Uniwersytet Warszawski, Warszawa 2018, s. 185–205. oraz D. Roszko, R. Roszko, *Korpusy wielojęzyczne wkładem Instytutu Sławistyki Polskiej Akademii Nauk w rozwój infrastruktury CLARIN-PL: Przykłady analizy korpusowej nad wołaczem*, w: J. L. Banasiak, A. Kiklewicz, J. Mazurkiewicz-Sułkowska (red.), *Języki słowiańskie dziś – w kręgu kategorii, struktur i procesów*, Instytut Sławistyki PAN – Wydawnictwo Uniwersytetu Łódzkiego, Warszawa – Łódź 2021, s. 281–313.

ślonością-nieokreślonością, modalnością, aspektem i temporalnością w językach polskim i litewskim. Ten udany eksperyment, potwierdzony wieloma publikacjami, doczekał się rozwinięcia w *Eksperymentalnym bułgarsko-polsko-litewskim korpusie (Bulgarian–Polish–Lithuanian Corpus*²) tworzonym praktycznie przez czteroosobowy polsko-bułgarski zespół w składzie: Ludmila Dimitrova, Violetta Koseska-Toszeva, Danuta Roszko i Roman Roszko. Oba te zasoby nigdy nie doczekały się publikacji ze względu na brak stosownych licencji. Jednak ich część została włączona do nowo powstających polsko-bułgarsko-rosyjskiego i polsko-litewskiego korpusów, konstruowanych przez Zespół Semantyki i Lingwistyki Korpusowej IS PAN (ZSiLK) w strukturach CLARIN-PL. Polskie Konsorcjum CLARIN-PL³ tworzą zespoły badawcze sześciu instytucji naukowych: Instytutu Podstaw Informatyki PAN, Instytutu Sławistyki PAN, Politechniki Wrocławskiej (lider konsorcjum), Polsko-Japońskiej Akademii Technik Komputerowych, Uniwersytetu Łódzkiego oraz Uniwersytetu Wrocławskiego. Celem Konsorcjum CLARIN-PL jest utworzenie polskiej infrastruktury badawczej na rzecz rozwoju nauk humanistycznych i społecznych w Polsce w tych obszarach badawczych, których podstawą jest analiza danych językowych. W ramach wsparcia badaczy szeroko rozumianych nauk humanistycznych i społecznych Konsorcjum Clarin-PL utworzyło spójną infrastrukturę, którą nieustannie rozwija i dostosowuje do zmieniających się standardów światowych. Jest to otwarta sieć, dostępna dla wszystkich badaczy, zapewniająca prowadzenie badań z wykorzystaniem nowoczesnych metod opartych na nieprzerwanie udoskonalanych technologiach przetwarzania języka naturalnego. Finansowanie działalności Konsorcjum CLARIN-PL rozpoczęło się w roku 2013. W latach 2013–2021 Konsorcjum zrealizowało trzy duże granty dotowane przez Ministerstwo Edukacji i Nauki, tworząc wiele narzędzi i zasobów językowych. Na początku 2020 roku Konsorcjum CLARIN-PL uzyskało kolejne dofinansowanie w Programie Operacyjnym Inteligentny Rozwój 2014–2020 (Priory-

² Szerzej na ten temat: L. Dimitrova, V. Koseska-Toszeva, D. Roszko, R. Roszko, *Bulgarian–Polish–Lithuanian Corpus: Current development*, w: C. Vertan, S. Piperidis, E. Paskaleva, M. Slavcheva (red.), *International Workshop: Multilingual resources, technologies and evaluation for Central and Eastern European languages held in conjunction with the International Conference RANLP–2009: PROCEEDINGS*, Borovets 2009, s. 1–8. L. Dimitrova, V. Koseska-Toszeva, D. Roszko, R. Roszko, *Trilingual Aligned Corpus: Current state and new applications*, „Cognitive Studies / Études cognitives” 2014, vol. 14, s. 13–20.

³ <http://clarin-pl.eu/> (11.11.2021).

tet IV: Zwiększenie potencjału naukowo-badawczego, Działanie 4.2: Rozwój nowoczesnej infrastruktury badawczej sektora nauki) na realizację nowych zadań.

W odniesieniu do tytułowych rusycystycznych akcentów w infrastrukturze CLARIN-PL należy przede wszystkim wymienić ukończone korpusy równoległe: polsko-bułgarsko-rosyjski, polsko-rosyjski (v.1 i v.2), rosyjsko-bułgarski, rosyjsko-litewski i rosyjsko-ukraiński oraz konstruowany obecnie zupełnie nowy polsko-rosyjski zasób w ramach szerszego zadania budowy ręcznie zrównoleglonych i znakowanych dwujęzycznych korpusów równoległych.

2. POLISH–BULGARIAN–RUSSIAN PARALLEL CORPUS

W latach 2013–2016 ZSiLK w składzie Anna Kisiel (do IX 2015), Violetta Koseska-Toszewa, Natalia Kotsyba, Joanna Satoła-Staškowiak i Wojciech Sosnowski pracowali nad bazą tekstów współczesnych w językach polskim, bułgarskim i rosyjskim. Wynikiem ich prac jest zasób, któremu nadano nazwę *Polish–Bulgarian–Russian Parallel Corpus*⁴. Objętość tego korpusu wyniosła 6 479 367 słowoform. Łącznie zawiera on 162 jednostki (są to pojedyncze utwory lub większe zbiory), w tym 78 plików to utwory beletrystyczne, 72 – teksty unijne reprezentujące język prawny, 6 – artykuły naukowe, po 3 – teksty prawnicze i religijne. Początkowy zamiar utworzenia korpusu, który składałby się tylko z utworów wzajemnie tłumaczonych (w ramach tej trójki języków) nie został osiągnięty. Ostatecznie tylko 24 jednostki bazują na wzajemnych tłumaczeniach, pozostałe w liczbie 138 są tłumaczeniami z języków trzecich (głównie z języka angielskiego). Zbiory będące wynikiem przeprowadzonych prac zostały opatrzone spójnym opisem metadanych w formacie CMDI i opublikowane w cyfrowym systemie repozytoryjnym *dSpace*⁵. Każdy użytkownik infrastruktury CLARIN-PL może pobrać wszystkie lub tylko wybrane pozycje z tej bazy w formacie TMX⁶.

⁴ A. Kisiel, V. Koseska-Toszewa, N. Kotsyba, J. Satoła-Staškowiak, W. Sosnowski, 2016, *Polish-Bulgarian-Russian Parallel Corpus*, CLARIN-PL digital repository 2016, <http://hdl.handle.net/11321/308> (11.11.2021).

⁵ <https://clarin-pl.eu/dspace/handle/11321/308> (11.11.2021).

⁶ TMX (< ang. *Translation Memory eXchange* / pol. *wymiana pamięci tłumaczeniowej*) – jest to jeden z najlepiej rozpoznawalnych standardów zapisu plików wymiany pamięci tłumaczeniowej (TM < ang. *Translation Memory* / pol. *pamięć tłumaczeniowa*).

3. POLISH–RUSSIAN PARALLEL CORPUS (2018)

Po reorganizacji ZSiLK w okresie 2016–2018 skonstruowano pięć dwujęzycznych korpusów równoległych z językiem polskim jako węzłowym, w tym polsko-rosyjski: *Polish–Russian Parallel Corpus*⁷. Zasób ten został opublikowany na stronach CLARIN-PL jednocześnie w cyfrowym systemie repozytoryjnym *dSpace*⁸ oraz w wielofunkcyjnym narzędziu *KonText*⁹. Czym jest wspomniany *KonText*? Jest to zaawansowane i nieustannie modernizowane narzędzie, oferujące użytkownikowi zarówno tworzenie własnych (dla języka polskiego automatycznie znakowanych) korpusów, jak i przeszukiwanie wielu udostępnionych w nim zasobów jedno- i wielojęzycznych. Sposoby przeszukiwania zbiorów są bardzo rozbudowane, a zastosowany w zapytaniu język CQL¹⁰ obsługuje atrybuty, operatory, wyrażenia regularne, lematy, tagi, klasy i fleksemy słów, kategorie gramatyczne i metaanotację. Interfejs *KonTextu* jest przejrzysty. Początkujący użytkownik bez problemu uzyska interesujące go profilowanie (sortowanie, filtrowanie) wyszukiwanych form, wyrażeń, konstrukcji oraz konkordancję zarówno w lewym jak i prawym kontekście od wybranej formy szukanej¹¹. *KonText* umożliwia automatyczne obliczanie wielu miar. Na stronach Clarin-PL jest dostępna instrukcja wyszukiwarki korpusowej *KonText*¹² oraz tzw. ściągawka instrukcji stosowanych w *KonText(-cie)*¹³.

W odniesieniu do opisanej w punkcie 2. wielojęzycznej bazy ten polsko-rosyjski korpus jest jej pewnym rozwinięciem dostosowanym do wymogu zbioru dwujęzycznego. W pracach nad jego powstaniem

⁷ R. Roszko, W. Sosnowski, M. Duszkin, D. Roszko, R. Tymoshuk, *Polish-Russian Parallel Corpus*, CLARIN-PL digital repository 2018, <http://hdl.handle.net/11321/534> (11.11.2021).

⁸ <https://clarin-pl.eu/dspace/handle/11321/534> (11.11.2021).

⁹ Adres wyszukiwarki: https://kontext.clarin-pl.eu/run.cgi/first_form (11.11.2021). Opis *KonText-u*: T. Machálek, *KonText: Advanced and flexible corpus query interface*, w: *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, European Language Resources Association 2020, s. 7003–7008.

¹⁰ CQL (< ang. *Corpus Query Language / pol. język zapytań korpusowych*) — jest to stworzony z myślą o efektywnym przeszukiwaniu zasobów korpusowych język zapytań.

¹¹ KWIC (< ang. *Key Word In Context / pol. słowo kluczowe/węzłowe w kontekście*).

¹² <https://nextcloud.clarin-pl.eu/index.php/s/fzAZg9xbxA4YEdu> (11.11.2021).

¹³ <https://nextcloud.clarin-pl.eu/index.php/s/IsIriR9v5Hopaml#pdfviewer> (11.11.2021).

uczestniczył ZSiLK w składzie: Maksim Duszkin, Danuta Roszko, Roman Roszko, Wojciech Sosnowski i Roman Tymoshuk. Objętość wyniosła 5 615 274 słowoforn. Całość korpusu liczy 152 jednostki. Zrównoważenie zasobów względem bazy wyjściowej zostało zmienione. Dołączono nowe teksty reprezentujące rejestr zbliżony do języka mówionego, mianowicie dialogi filmowe w liczbie 101. Język prawny jest reprezentowany przez 26 jednostek, beletrystyka – 12. Pozostałe pliki stanowią teksty naukowe, religijne i techniczne (razem 13). Zasoby nie zostały oznakowane morfologicznie.

Opublikowane w roku 2018 równoległe korpusy CLARIN-PL zostały zastosowane przez wielu badaczy¹⁴ i zespoły badawcze (polsko-bułgarski, polsko-bułgarsko-ukraiński)¹⁵. Znane są zastosowania tych zasobów w pracach licencjackich (Katedra Językoznawstwa Ogólnego, Migowego i Bałtystyki UW), magisterskich (Instytut Lingwistyki Stosowanej UW, Katedra Językoznawstwa Ogólnego, Migowego i Bałtystyki UW, Wydział Polonistyki UW), doktorskich (UW, UJ) i habilitacyjnych (Instytut Sławistyki PAN).

4. POLISH–RUSSIAN PARALLEL CORPUS „2” (2022) ORAZ INNE NOWE KORPUSY DWUJĘZYCZNE Z JĘZYKIEM ROSYJSKIM (2022)

4.1. POLISH–RUSSIAN PARALLEL CORPUS „2”

W latach 2018–2021 opisany w punkcie 3. korpus podlegał dalszemu rozwojowi. Zadania były realizowane przez ZSiLK w składzie: Maksim Duszkin, Danuta Roszko, Roman Roszko we współpracy z Jewgieniją Żejmo¹⁶. W tym okresie zwrócono szczególną uwagę na wstępne prze-

¹⁴ Między innymi warto przywołać prace Jakuba Banasiaka nad semantyką, por. J. Banasiak, *Built-in argument positions in Bulgarian and Polish*. “Cognitive Studies / Études cognitives”, 2021(21).

¹⁵ Wypada wskazać zespoły międzyuczelniane i międzynarodowe prowadzące prace nad leksyką w językach słowiańskich, por. W. Sosnowski, J. Satoła-Staškowiak, *A contrastive analysis of feminines in Bulgarian, Polish and Russian*. “Cognitive Studies / Études cognitives” 2019 (19), Article 1922; D. Blagoeva, M. Jaskot, W. Sosnowski, *A lexicographical approach to the contrastive analysis of Bulgarian and Polish phraseology*, „Cognitive Studies / Études cognitives” 2019(19), Article 1923.

¹⁶ Szerzej na ten temat por. M. Duszkin, D. Roszko, R. Roszko, *New parallel corpora of Baltic and Slavic languages – Assumptions of corpus construction*, w: K. Ekštejn, F. Pártl, M. Konopík (red.), *Lecture Notes in Artificial Intelligence LNAI 12848: Text, Speech, and Dialogue TSD 2021*, Springer International Publishing, Cham 2021, s. 173–183.

tworzenie już gotowych zasobów, w którego procesie zlikwidowano wiele błędów pisowni, uzgodniono wersje tłumaczeń oraz wyróżniono najmniejsze jednostki (tzw. tokeny). Ponadto w odpowiedzi na propozycje użytkowników CLARIN-PL zmodyfikowano tak normy zrównoleglenia zasobów, by poszczególne wydzielane jednostki były możliwie najmniejsze. Przyjęta na wcześniejszych etapach projektowych zasada jednoczesnego wiązania dwóch równoległych tekstów na poziomie segmentów zdaniowych spełniającego warunek zupełności komunikatywnej, doskonale sprawdzająca się w opisie jednego języka, została rozwinięta. We wcześniejszej wersji korpusów CLARIN-PL przyjmowano, że zarówno początek zdania, jak i jego koniec mają następujące formalne wykładniki: wielką literę (jako początek zdania) i kropkę lub wykrzyknik, lub pytajnik, lub wielokropkę (jako znak kończący zdanie). Ze względu na ograniczoną objętość artykułu pomijam dodatkowe warunki, które muszą zostać spełnione, by te elementy formalne mogły zostać uznane za początek lub koniec zdania. Zgodnie z tymi założeniami każdy ciąg rozpoczęty wielką literą i zakończony jednym ze wspomnianych znaków końca był uznawany za zdanie, któremu przypisywano zupełność komunikatywną. Jednoczesna segmentacja na poziomie zdania — jak wiadomo — wiązała się ze wskazaniem najmniejszych równoległych partii tekstu, dla których byłby spełniony warunek jednoczesnego zachodzenia w obu językach początku i końca zdania oraz zachowana zgodność treści odczytywanych na płaszczyźnie semantycznej. W rezultacie takich założeń uzgodnione międzyjęzykowe segmenty mogły zawierać nawet po kilka-kilkanaście zdań. W przypadku niektórych utworów beletrystycznych faktycznie cały utwór wypełniał znamiona jednego segmentu zdaniowego, na przykład opowiadanie *Miasto laikrodis* litewskiego pisarza Algisa Kuklysa jest jednym długim zdaniem liczącym około 1500 wyrazów. W związku z takim stanem rzeczy analizowano równoległe zasoby wielojęzyczne i starano się znaleźć inne formalnie wyrażone charakterystyczne miejsca, które warunkowo można uznać za koniec zdania. Wśród nowych znaczników, które w ściśle określonych warunkach mogą być potraktowane jako wskaźnik końca zdania, znalazły się następujące znaki interpunkcyjne: przecinek i nawiasy (szczególnie w tekstach prawnych), dwukropki i średniki (te głównie w utworach beletrystycznych), por. tab. 1–2, a szczególnie w tab. 2 korespondujące ze sobą w rosyjskiej i bułgarskiej wersji tekstu różne znaki przestankowe kropki i średnika.

Tabela 1. Propozycja rozbicia segmentu na kilka mniejszych z wykorzystaniem znaków interpunkcyjnych «;» i «,» w utworze beletrystycznym

Язык русский	Язык български
<i>Дождь был косой, ниспадала хлещущая сплошная завеса;</i>	<i>Дъждът плющеше и полегатите му струи ту се засилваха, ту отслабваха;</i>
<i>изредка перед глазами Джима вставали грозно надвигающиеся волны,</i>	<i>от време на време пред очите на Джим се появяваха страшно надигащи се вълни;</i>
<i>маленькое суденышко металось у берега;</i>	<i>малкото корабче се мяташе край брега;</i>
<i>неподвижные строения вырисовывались в плавучем тумане;</i>	<i>неподвижните постройки се открояваха сред плаващата мъгла;</i>
<i>тяжело раскачивались широкие паромы на якорь,</i>	<i>тежко се люшкаха закотвените широки фериботи;</i>
<i>поднимались и опускались огромные пристани, задушенные брызгами.</i>	<i>издигаха се и се спускаха обширните кейове, задавени от пръски.</i>

Tabela 2. Propozycja rozbicia segmentu na kilka mniejszych z wykorzystaniem znaków interpunkcyjnych «;» w języku rosyjskim, formalnie odpowiadających «.» w bułgarskim (utwór beletrystyczny).

Язык русский	Язык български
Глаза твои голубиные под кудрями твоими;	Очите ти са гълъбови под твоите къдри.
волосы твои — как стадо коз, сходящих с горы Галаадской;	Косата ти е като стадо кози, кога слиза от Галаатската планина.
зубы твои — как стадо выстриженных овец, выходящих из купальни, из которых у каждой пара ягнят, и бесплодной нет между ними;	Зъбите ти като стадо овце, кога излиза от къпалня, от които всяка с по две агънца, и ялова няма помежду им.

W tekstach prawnych i prawniczych zdecydowano się na uwzględnienie w segmentacji miejsc wyznaczonych strukturą dokumentu, por. tab. 3.

Tabela 3. Propozycja rozbicia zdania na segmenty
zgodne ze strukturą dokumentu prawnego (języki polski i litewski)

Język polski	Język litewski
Dziennik Urzędowy Unii Europejskiej	Europos Sąjungos oficialusis leidinys
C 336 / 19	C 336 / 19
z dnia 28 czerwca 2011 r.	2011 m. birželio 28 d.
w sprawie E-18 / 10	byloje E-18 / 10
Urząd Nadzoru EFTA przeciwko Królestwu Norwegii	ELPA priežiūros institucija prieš Norvegijos Karalystę
(Niewykonanie wyroku Trybunału stwierdzającego uchybienie zobowiązaniom – artykuł 33 porozumienia o nadzorze i Trybunale – środki niezbędne do wykonania wyroku Trybunału)	(Įsipareigojimų neįvykdymą konstatuojančio Teismo sprendimo neįvykdymas – Institucijos ir Teismo susitarimo 33 straipsnis – Priemonės Teismo sprendimui įvykdyti)
2011 / C 336 / 13	2011 / C 336 / 13

Zgodnie z nowo opracowanymi zasadami segmentacji na poziomie zdania dokonano ręcznej korekty/aktualizacji zrównoleglenia zasobów. Następnie wszystkie tokeny oznakowano w automatycznym procesie. Całość polskich tekstów otagowano narzędziem *MorphoDiTa-PL*¹⁷, rosyjskie zaś – *UDPipe (russian-gsd-ud-2.6-200830)*¹⁸, por. przykład [1]. Poszczególne jednostki opatrzone takimi metadanymi jak: *text.tdtype* (zaszeregowanie tekstu do odpowiedniego rejestru), *text.srclang* (język oryginału), *text.original* (oznaczenie rozróżniające teksty oryginalne i tłumaczenia, przewidziano również wartość "unknown"), *text.author* (autor tekstu), *text.translator* (autor tłumaczenia), *text.srctitle* (nazwa tekstu w języku tekstu pierwotnego), *text.title* (nazwa utworu w danym języku), *text.year* (data publikacji utworu), *text.period* (zaszeregowanie utworu do wyróżnionych okresów), *text.url* (adres internetowy), por. przykład [2].

¹⁷ <https://ws.clarin-pl.eu/morphoDiTa.shtml> (11.11.2021). M. Piasecki, W. Walentynowicz, *MorphoDiTa-based tagger adapted to the Polish language technology*, w: *Proceedings of Human Language Technologies as a Challenge for Computer Science and Linguistics*, LTC 2017, Poznań 2017 s. 377–381.

¹⁸ <http://lindat.mff.cuni.cz/services/udpipe/> (11.11.2021). M. Straka, J. Straková, *UDPipe*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, Prag 2016, <http://hdl.handle.net/11234/1-1702> (11.11.2021).

[1] Oznakowanie segmentu 4 w języku rosyjskim: *Международный инструментальный конкурс «ЗОЛОТЫЕ КЛАВИШИ» — ONLINE*¹⁹.

sent_id = 4

1	Международный	международный	ADJ	JJL	C a s e -			
	=Nom/Degree=Pos/Gender=Masc/Number=Sing	_	_	_	J J L	_	_	_
2	инструментальный	инструментальный	ADJ	J J L				
	Case=Nom/Degree=Pos/Gender=Masc/Number=Sing	_	_	_				
3	конкурс конкурс	NOUN	NN	Animacy=Inan/Case=Nom/Gender=Masc/Number=Sing	_	_	_	_
4	«ЗОЛОТЫЕ	«золотой	ADJ	JJL	Case=Nom/Degree=Pos/Number=Plur	_	_	_
5	КЛАВИШИ»	клавиши»	NOUN	NN	Animacy=Inan/Case=Nom/Gender=Fem/Number=Plur	_	_	_
6	-	-	PUNCT	(
7	ONLINEONLINEX	FW	Foreign=Yes					

Widoczne w przykładzie [1] zapisy nie są widoczne dla użytkownika, jednak mają niebagatelne znaczenie podczas przeszukiwania korpusu. Wszystkie przypisane formom parametry mogą być częścią zapytania. Spójrzmy na 4. słowoformę (*ЗОЛОТЫЕ*) w przykładzie [1]. W jej opisie pojawia się lemat *золотой* zgodny z formą hasłową w słowniku, następnie po nim następują tagi, które wskazują na wartości morfologiczne przypisane tej słowoformie (*ЗОЛОТЫЕ*) w danym kontekście: część mowy = przymiotnik / przypadek = mianownik / stopień = równy / liczba = mnoga.

Poniżej zamieszczony zostaje plik informacyjny z metadanymi opisującymi tekst zawierający przedstawione w przykładzie [1] zdanie nr 4 (# sent_id = 4).

[2] Przykładowy plik z metadanymi do tekstu rosyjskojęzycznego.

```
text.tstype"journalism"
text.srclang"bg Bulgarian"
text.original"no"
text.author"European Association of Folklore Festivals"
text.translator"European Association of Folklore Festivals"
text.srctitle"ЗЛАТНИ КЛАВИШИ"
text.title"ЗОЛОТЫЕ КЛАВИШИ"
```

¹⁹ W pozycji 5. zastosowany przez nas do znakowania zasobów rosyjskich program *UDPipe* (russian-gsd-ud-2.6-200830) wygenerował lemat *клавиши* w formie liczby mnogiej zamiast spodziewanego *клавиша*. Wariant lematu w liczbie pojedynczej jest generowany przez *UDPipe* (russian-taiga-2.6-200830). Wybór zastosowanego modelu do znakowania był poprzedzony testami, które wskazywały na statystycznie wyższą trafność russian-gsd-ud-2.6-200830.

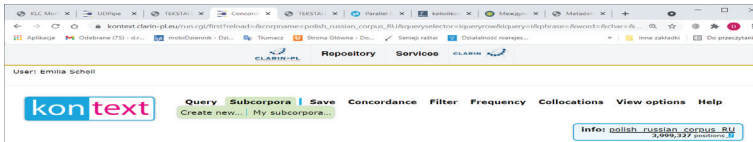
text.year"2021"

text.period"after 1990"

text.url"https://eaff.eu/ru/championships/1011-2811-international-instrumental-competition-golden-keys-online"

Parametry zawarte w metadanych również mogą być częścią zapytania. Na przykład użytkownik zajmujący się użyciem przymiotników po roku 1990 w języku rosyjskim powinien skorzystać z dwu możliwości zawężenia przeszukiwanych plików do datowanych na rok 1990 i późniejsze. Pierwszym sposobem jest utworzenie własnego podkorpusu składającego się z plików spełniających wybraną wartość "after 1990" dla parametru text.period, por. przykład [3], drugim zaś — każdorazowe zaznaczanie poniżej pola zapytania w części text.period wartości "after 1990".

[3] KonText: Menu tworzenia i wyboru własnych podkorpusów.



Powracając do badań nad rosyjskim przymiotnikiem w okresie po 1990 roku użytkownik ponadto winien w polu wpisywania zapytania zamieścić właściwą formułę, por. przykład [4].

[4] Zapytanie o wyświetlenie wszystkich zdań/segmentów zawierających przymiotnik w zasobach rosyjskojęzycznych.
[tag="ADJ.*"]

W przykładzie [4] nawiasy kwadratowe wyznaczają granicę słowofomy, której przypisano jeden parametr *tag* o wartości *ADJ* (przymiotnik). Następujące po *ADJ* znaki *.** oznaczają, że inne doprecyzowujące wartości są obojętne. Jak można zauważyć, wartość parametru (w tym przypadku *tag*) jest umieszczana w cudzysłowie (tu: "ADJ.*").

W poniższym przykładzie [5] zostaje podana składnia zawężająca parametry wyszukiwania. Załóżmy, że badacz jest zainteresowany użyciem przymiotników rosyjskich rozpoczynających się tylko na literę *э*. W tym celu rozbudowuje formułę o parametr *lexem* (lemat), któremu w cudzysłowie przypisuje wartość *э.** (tj. "э.*") i dodaje operator koniunkcji *&*, by obie wartości wymienionych parametrów zachodziły jednocześnie.

[5] Rozbudowa zapytania z przykładu [4] zawężająca wynik do zdań/segmentów zawierających przymiotnik rozpoczynający się od litery э (np. *этнологический, экономические, эстетического* itd.) w języku rosyjskim.
[lexem="э.*" & tag="ADJ.*"]

W przykładzie [6] zostaje przedstawione zapytanie o dwuwyrzowe wyrażenia złożone zawierające przymiotnik rozpoczynający się od litery э, po którym bezpośrednio następuje rzeczownik rozpoczynający się od litery ф.

[6] Rozbudowa zapytania z przykładu [5] zawężająca wynik do zdań/segmentów zawierających dwuwyrzowe złożenia przymiotnika (rozpoczynającego się od litery э) i rzeczownika (rozpoczynającego się od litery ф) (np. *этнологического факультета, экономический форум, (на) эстетическом фоне* itd.) w języku rosyjskim.
[lexem="э.*" & tag="ADJ.*"] [lexem="ф.*" & tag="NOUN.*"]

Wspomniane w tej części artykułu wstępne przetworzenie zasobów objęło również wiele kolejnych istotnych dla użytkownika aspektów. Wprowadzono znaki umowne, które informują badacza o braku ekwiwalentu w jednym z języków korpusu. W tym celu zastosowano skrót [—], por. przykład [7]. Kolejny wprowadzony znak [...] informuje o skróceniu tekstu o nieistotne z punktu widzenia wielojęzycznej analizy fragmenty, por. przykład [8]. Ten drugi znak początkowo stosowano w celu ukrycia danych wrażliwych. W obecnie opracowywanych korpusach (o których poniżej w punkcie 5.) odstąpiono od tak przeprowadzanej anonimizacji na rzecz inteligentnej anonimizacji, która w pełni zachowuje strukturę tekstu, por. przykład [9].

[7] Użycie znaku umownego [—] na oznaczenie braku w którymkolwiek języku ekwiwalentnego segmentu.

PL	RU
Chcesz w pysk, to chodź!	Тебе нужны неприятности?
[—]	Сейчас они будут!
Podoba ci się?	Ну что, доволен?

[8] Użycia znaku umownego [...] na oznaczenie pominiętych fragmentów na etapie wstępnego przetworzenia (przykład polski).

Tekst na wejściu	Tekst po przetworzeniu
REPUBLIKA LITEWSKA	REPUBLIKA LITEWSKA
LT	LT

Instytucja Państwowa „Regitra”	Instytucja Państwowa „Regitra”
DOWÓD REJESTRACYJNY	DOWÓD REJESTRACYJNY
Свидетелство за регистрация / Permiso de circulation / Osvědčení o registraci / Registreringsattest / Zulassungsbescheinigung / Rcgistreerimistunnistus / Πιστοποιητικό εγγραφής / Registration certificate / Certificat d'immatriculation / Prometna dozvola / Teastas Claráithe / Carta di circolazione / Reģistrācijas apliecība / Forgalmi engedély / Čertifikat ta' Reġistrazzjoni / Kentekenbewijs / Dowód Rejestracyjny / Certificado de matrícula / Certificat de Înregistrare / Osvedčenie o evidencii / Prometno dovoljenje / Rekisteröintitodistus / Registreringsbeviset	[...]

[9] Inteligentna anonimizacja w języku polskim.

Tekst na wejściu	Tekst ze zidentyfikowanymi nazwami wrażliwymi	Tekst po przetworzeniu (po inteligentnej anonimizacji)
13 grudnia 1982 roku w Warszawie w Alejach Jerozolimskich został zatrzymany uczestnik strajku Jan Kowalczyk z narzeczoną Joanną Nowak.	[DATE] roku w [MIEJSCE] w [MIEJSCE] został zatrzymany uczestnik strajku [OSOBA] z narzeczoną [OSOBA].	21 grudnia 1900 roku w Arkadii na ulicy Cichej został zatrzymany uczestnik strajku Ananiasz Baran z narzeczoną Elżbietą Jerzy.

W okresie 2018–2021 obok prac nad dopracowaniem już zebranych zasobów, podjęta została rozbudowa korpusu polsko-rosyjskiego o kolejne teksty współczesne. Zgodnie z sugestiami dotychczasowych i potencjalnych użytkowników głównie skupiono się na pozyskaniu nowych zasobów reprezentujących rejestr jak najbardziej zbliżony do mowy, czyli na dialogach filmowych, oraz na dokumentach prawnych i naukowych. Szczególną uwagę zwrócono na jakość i wzajemną międzyjęzykową adekwatność części nieautoryzowanych tłumaczeń

list dialogowych. Podczas sprawdzania zrównoleglenia korektor wychwytywał niezgodności między obiema wersjami językowymi (na poziomie segmentów zdaniowych i leksykalnym) oraz niekonsekwencje tłumaczenia w każdym języku niezależnie²⁰. Na przykład występujący w wersji polskiej *plaszcz przeciwdeszczowy* w rosyjskiej wersji listy dialogowej raz był nazwany *шпнель* innym razem — *пальто*. Należy zauważyć, że w wersji rosyjskiej dochodzi do nazywania tego samego przedmiotu dwoma różnymi leksemami, które w żadnym stopniu nie są leksykalnymi ekwiwalentami polskiego *plaszcz przeciwdeszczowy*. Podobną rozbieżność na poziomie leksykalnym odnotowujemy w formach polskiej *kamizelka kuloodporna* i rosyjskiej *свитерок*. Dużo uwagi poświęcono ujednoczeniu zapisu nazw własnych, na przykład w wersji polskiej trzy formy imienia *Brian, Brajan, Brajen* zastąpiono jedną *Brian*, podobnie leksemy *Samanta* i *Samantha* zastąpiono jednym *Samantha*. W przypadku podwójnego zapisu *izraelici* i *Izraelici* wskazano postać *Izraelici* jako zgodną z normami języka polskiego. Ponadto korygowano rodzaj gramatyczny zaimków osobowych, by odpowiadał płci bohatera, wyrównywano rozbieżności w użyciu liczby pojedynczej i mnogiej²¹. W przykładzie [10] zebrano kilka egzemplifikacji rozbieżności formalnych i merytorycznych między językami polskim i rosyjskim oraz przedstawiono decyzję korektorów-tłumaczy. W wielu niejednoznacznych kontekstowo zapisach sprawdzano trafność tłumaczenia z wersją oryginalną listy dialogowej.

[10] Przykłady różnych strategii tłumaczeniowych w językach polskim i rosyjskim oraz (nie)wprowadzone zmiany w polsko-rosyjskich zasobach.

	PL	RU	Decyzja korektora
1	Nasza matka nas nie urodziła. // Zrobiła to taka inna pani.	Наша мама — не наша родная мама. // Другая женщина — наша родная мама.	Pozostawiono bez zmian.
2	Zostało mi jeszcze na jedną.	У меня хватит индейки еще на один сэндвич.	Pozostawiono bez zmian.
3	Stacja metra <i>State and Balboa</i> .	Станция метро, <i>переход со Стейт на Балбоа</i> .	Pozostawiono bez zmian ze względu na odmienne realia polskie i rosyjskie związane z infrastrukturą metra.

²⁰ Błędy pisowni zostały wyeliminowane podczas wstępnego przetwarzania. Błędy gramatyczne (np. użycie niewłaściwego przypadku), niektóre błędy interpunkcyjne były korygowane przez korektora-tłumacza.

²¹ W języku rosyjskim często należało zaimek *они* zamieniać na *он* lub *она*.

4	Ona też w tym siedzi.	Ты был прав по поводу этого дерьма.	Pozostawiono bez zmian.
5	Tego chcesz?	Нет.	Utworzono dwa segmenty, dodając dwukrotnie znak [–]: <i>Tego chcesz? / [–] [–] / Hem.</i>
6	Co tu się dzieje?	Что со мной происходит?	Zmiana zdania rosyjskiego jako wynik sprawdzenia oryginału: <i>Co tu się dzieje? / Что здесь происходит?</i>
7	Proszę, <i>podejź.</i>	Пожалуйста, <i>проходи.</i>	Zmiana zdania rosyjskiego jako wynik sprawdzenia oryginału: <i>Proszę, podejź. / Пожалуйста, подойди.</i>
8	Musiał użyć <i>defibrylatora.</i>	Врач был вынужден применить <i>электрошокер.</i>	Zmiana zdania rosyjskiego jako wynik sprawdzenia oryginału: <i>Musiał użyć defibrylatora. / Ему пришлось воспользоваться дефибрилятором.</i>
9	Ty <i>pierwszy.</i>	Ты <i>первая.</i>	Zmiana zdania rosyjskiego powodowana koniecznością zachowania zgodności odniesienia do współmówcy-mężczyzny: <i>Ty pierwszy. / Ты первый.</i>
10	<i>Marzyłem o tobie.</i>	Ты мне часто <i>снилась.</i>	Zmiana treści zdania rosyjskiego na bardziej odpowiadającą treści filmu. W angielskim oryginale został użyty leksem <i>dream</i> : <i>Marzyłem o tobie. / Я мечтал о тебе.</i>
11	Płaskie czy na <i>obcasach?</i> (buty)	Плоские или на <i>платформе?</i>	Zmiana leksemu w zdaniu rosyjskim po obejrzeniu sceny: <i>Płaskie czy na obcasach? / Плоские или на каблучках?</i>

RUSYCYSTYCZNE ZASOBY I NARZĘDZIA CLARIN-PL

12	<i>Idź, kochanie.</i>	<i>Продолжай, дорогая.</i>	Zmiana treści zdania rosyjskiego, by ją uzgodnić z kontekstem: <i>Idź, kochanie. / Иди, дорогая.</i>
13	Dlaczego?	Ты только приехал, что не так?	Zdecydowano się na wprowadzenie znaku [–] w języku polskim: [–] <i>Dlaczego? / Ты только приехал, что не так?</i>
14	Chcę, byś włożył kurczaka do piekarnika i ustawił na 375.	Мне нужно, чтобы ты поставила курицу в духовку на 375 градусов.	Zmiana wskazań temperatury w obu językach, by jej wartość została właściwie odczytana przez polskich i rosyjskich odbiorców (zastosowano wskazania w stopniach Celsjusza): <i>Chcę, byś włożył kurczaka do piekarnika i ustawił na 190. / Мне нужно, чтобы ты поставила курицу в духовку на 190 градусов.</i>
15	Wirus po 24 godzinach doprowadza do zgonu.	Вирус станет инертным через 24 часа.	Zmiana wprowadzona w zdaniu polskim, by odpowiadała realiom przedstawianym w filmie: <i>Wirus staje się obojętny po 24 godzinach. / Вирус станет инертным через 24 часа.</i>
16	<i>Być czym?</i>	<i>Где?</i>	Usunięto niepowiązane z kontekstem segmenty (nie znaleziono oryginału tej listy).

[11] Przykład spektakularnej zmiany dokonanej przez korektora-tłumacza w zasobach polsko-ukraińskich.

PL	UK
Wersja zgodna z zapisem w liście dialogowej	
Byliśmy na wycieczce w Universal Studio, a potem w woskowym muzeum Hollywood.	Ми ходили на екскурсію по кіностудії Довженка, а потім Чарлі водив мене у музей воскових фігур Верховної Ради.

Wersja po zmianach korektora-tłumacza ²³	
Byliśmy na wycieczce w Universal Studios, a potem w Muzeum Figur Woskowych w Hollywood.	Ми ходили на екскурсію по <i>Студії Universal</i> , а потім Чарлі водив мене в <i>Голлівудський музей воскових фігур</i> .

Korekta list dialogowych (głównie polsko-rosyjskich i sporadycznie polsko-ukraińskich) dotyczyła tylko części z nich²². Autoryzowane przekłady nie podlegały jakimkolwiek zmianom. Informacja o wprowadzeniu tego typu poprawek w wybranych tekstach została zawarta w skorelowanych z nimi plikach metadanych. Decyzję o korekcie części list dialogowych podjęto z myślą o uczeniu maszynowym. Otóż złe dane wejściowe prowadzą do wytworzenia błędnych międzyjęzykowych modeli, a te później zastosowane w translatorach generują błędne tłumaczenia.

4.2. KORPUSY ROSYJSKO-* -BUŁGARSKI, -LITEWSKI I -UKRAIŃSKI (2022)

Równoległe z pracami nad polsko-rosyjskim korpusem prowadzono konstrukcję dziesięciu nowych korpusów, w tym trzech z językiem rosyjskim: rosyjsko-litewskiego, rosyjsko-bułgarskiego i rosyjsko-ukraińskiego. Nie są to duże zasoby, zwłaszcza rosyjsko-litewskie. Zrównoważenie tych kolekcji jest różne. Najlepsze udało się osiągnąć dla pary rosyjsko-bułgarskiej, która zawiera 44 jednostek, w tym 27 reprezentujących rejestr prawny, 11 – beletrystykę (wśród nich jeden utwór literatury dziecięcej), 3 – rejestr prawniczy, 2 – rejestr tekstów naukowych, 1 – rejestr dziennikarski. Zrównoważenie części rosyjsko-ukraińskiej jest następujące: 25 jednostek reprezentujących język prawny, 3 – to teksty naukowe i 1 utwór beletrystyczny. Korpus rosyjsko-litewski zawiera tylko kilka tekstów prawniczych. Wszystkie te zasoby zostały wstępnie przetworzone, ręcznie zrównoleglone, automatycznie oznakowane i opatrzone metadanymi. Do

²² Podczas konstrukcji korpusów rozważaliśmy pominięcie niedoskonałych przekładów, jednak szczegółowa analiza tych „amatorskich” tłumaczeń dialogów filmowych wykazała użycie w nich wielu form, wyrażeń i konstrukcji typowych dla języka potocznego. Uznaliśmy więc, że warto niektóre z tych tekstów „uratować” poprzez naniesienie niezbędnej korekty językowej. Wprowadzane przez nas zmiany są wręcz konieczne w przypadku wykorzystania tych zasobów w uczeniu, testowaniu i konstrukcji nowych narzędzi lingwistycznych.

²³ Pochyleniem zaznaczono zmieniane fragmenty zdań w obu językach. Korekta w języku ukraińskim jest o wiele dalej idąca. Została ona podyktowana miejscem, w którym toczy się akcja filmu (Hollywood). Należy tu podkreślić, że przyjęta przez ukraińskiego tłumacza strategia jest niezrozumiała.

znakowania tekstów rosyjskojęzycznych zastosowano znane już narzędzie *UDPipe*¹⁸, dla litewskich – *KLC Morfologijos servisas*²⁴, bułgarskich – *CL@RK System*²⁵ i ukraińskich *UDPipe (ukrainian-ii-ud-2.6200830)*¹⁸.

W 2022 roku jest planowana publikacja tych korpusów na stronach CLARIN-PL w przeglądarce *KonText* oraz cyfrowym repozytorium *dSpace*.

5. RĘCZNIE ZRÓWNOLEGLONY I RĘCZNIE ZNAKOWANY POLSKO-ROSYJSKI KORPUS TEKSTÓW RÓWNOLEGLYCH

Od 2020 roku jest realizowany projekt CLARIN-PL-BIZ w ramach „Programu Operacyjnego Inteligentny Rozwój 2014–2020” w osi priorytetowej IV „Zwiększenie potencjału naukowo-badawczego” i działaniu 4.2 „Rozwój nowoczesnej infrastruktury badawczej sektora nauki”. Jego celem jest utworzenie platformy badawczo-rozwojowej do przetwarzania języka naturalnego i eksploracji dużych zasobów danych językowych. Obejmuje ona wiele zadań, w tym:

- utworzenie Centrum Technologicznego (CTech) jako bazy dla technologii eksploracji danych językowych;
- zastosowanie w CTech rozbudowanych technologii językowych do inteligentnego przetwarzania dużych różnorodnych danych na niewspieranych dotychczas płaszczyznach;
- opracowanie i wdrożenie właściwych standardów konstrukcji zasobów i narzędzi językowych;
- zaprojektowanie i wytworzenie nowych narzędzi analizy danych językowych;
- przygotowanie i dostarczenie nowych danych ręcznie znakowanych do badań i trenowania narzędzi językowych²⁶, a wśród nich również polsko-rosyjskiego korpusu²⁷.

²⁴ Przygotowując korpusy, uzyskaliśmy dostęp do najnowszej wersji tagera języka litewskiego *KLC Morfologijos servisas*, opracowanego przez Centrum Lingwistyki Komputerowej Uniwersytetu Witolda Wielkiego w Kownie. W momencie przygotowania tego artykułu wspomniany tager nie został jeszcze udostępniony.

²⁵ K. Simov, A. Simov, P. Osenova, *An XML architecture for shallow and deep processing*, w: *The Proceedings of the ESLLI 2004 Workshop on Combining Shallow and Deep Processing for NLP*, ESLLI 2004, s. 51–60.

²⁶ Szerzej na temat projektu por. R. Roszko, *O nowych ręcznie zrównoleglonych i znakowanych dwujęzycznych korpusach równoległych oraz ich zastosowaniach*, „Acta Baltico-Slavica” 2021, t. 45, artykuł 2576 oraz <https://clarin.biz/> (11.11.2021).

²⁷ Trwają prace nad czterema tego typu korpusami. Obok polsko-rosyjskiego tworzone są korpusy polsko-słoweński, polsko-litewski i polsko-bułgarski.

Odbiorcami ręcznie zrównoleglonych i znakowanych zasobów równoległych będą przede wszystkim programiści, informatycy zajmujący się projektowaniem i wytwarzaniem narzędzi do przetwarzania języka naturalnego (NLP)²⁸ oraz badacze reprezentujący szeroko rozumiane nauki humanistyczne i społeczne. Należy zwrócić uwagę, że ręcznie zrównoleglone i znakowane korpusy (w tym polsko-rosyjski) znajdują zastosowanie w definiowaniu algorytmów projekcji znaczeń z jednego języka na drugi. Szczególnie chodzi tu o projekcję tych znaczeń, które w jednym języku są jednoznacznie wyrażane na płaszczyźnie formalnej, w drugim zaś — dochodzi do tzw. niedopowiedzenia językowego²⁹. Jak wiadomo, ręcznie zrównoleglone i znakowane korpusy równoległe odgrywają dużą rolę w dopracowaniu narzędzi do automatycznego wyrównywania zasobów dwujęzycznych, co w konsekwencji przekłada się na wzrost liczby i jakości nowych kolekcji wielojęzycznych generowanych automatycznie. Nie można nie wspomnieć o znaczeniu tych zasobów w rozwoju rekurencyjnych sieci neuronowych, mających zastosowanie w przekładzie maszynowym i rozwoju sztucznej inteligencji.

5.1. ZASADY KONSTRUKCJI RĘCZNIE ZRÓWNOLEGLONEGO I ZNAKOWANEGO POLSKO-ROSYJSKIEGO KORPUSU RÓWNOLEGLĘGO

W realizowanym projekcie CLARIN-PL-BIZ na przełomie lat 2020/2021 przeprowadzono gruntowną analizę istniejących zasobów cyfrowych dla języków słowiańskich i bałtyckich³⁰, a następnie wypracowano metodologię konstrukcji poszczególnych ręcznie zrównoleglonych i znakowanych dwujęzycznych korpusów równoległych. Dużą wagę przywiązano do selekcji utworów, w tym oceny poprawności językowej tekstów oraz wewnętrznego zrównoważenia

²⁸ Przykłady narzędzi NLP: systemy do przechowywania i udostępniania danych językowych, wyszukiwarki korpusowe, analizatory cech gramatycznych, składniowych, stylometrycznych, anotatory znakujące/kodujące zasoby językowe (np. tagery, lematyzatory), syntezytory mowy, systemy do przetwarzania mowy itd.

²⁹ V. Koseska, R. Roszko, *On semantic annotation in CLARIN-PL parallel corpora*, "Cognitive Studies / Études cognitives" 2015, nr 15, s. 211–236. <https://doi.org/10.11649/cs.2015.016> (11.11.2021).

³⁰ Wyniki badań przedstawiono w czterech obszernych raportach CLARIN-PL-BIZ autorstwa Jakuba Banasiaka (IS PAN), Pawła Kowalskiego (IS PAN), Danuty Roszko (UW) i Romana Roszko (IS PAN). Szerzej na temat raportów por. R. Roszko, *O nowych ręcznie zrównoleglonych i znakowanych...*

zasobów. Przewidziano korektę pisowni, w przypadku danych wrażliwych także „inteligentną” anonimizację, por. wyżej przykład [9]. Segmentacja na poziomie zdania (zrównoleglanie/wyrównywanie) oraz tokenów (znakowanie/tagowanie) będzie przeprowadzana dwuetapowo. W pierwszym etapie dwa niezależne zespoły dokonają ręcznej segmentacji i znakowania. W drugim zaś — zespół superanotatorów w przypadkach rozbieżności wskaże właściwą segmentację i znakowanie. Wszystkie pliki włączone do zasobu będą posiadały rozbudowane metadane.

Rozważane jest pozyskiwanie quasi-równoległych danych w automatycznym przeszukiwaniu zasobów sieciowych z wykorzystaniem narzędzi LASER³¹ czy LAMBERT³². Decyzja o włączeniu tego typu danych do polsko-rosyjskiego korpusu jeszcze nie zapadła. Quasi-równoległe dane na pewno uzupełniłyby rejestr dziennikarski, który głównie tworzyłyby krótkie doniesienia prasowe oraz wypowiedzi znanych w świecie przywódców duchowych czy polityków. Należy jednak zważyć, że tego typu dane to przede wszystkim krótkie/hasłowe informacje prasowe o quasi-równoległym charakterze. Ponadto takie dane zamierza w tym projekcie gromadzić zespół prof. Piotra Pęczika, którego quasi-równoległe dane wraz z ręcznie zrównoleglanymi i znakowanymi dwujęzycznymi korpusami będą stanowić podstawę do budowy międzyjęzycznych modeli na rzecz maszynowego przekładu i dalszego rozwoju sztucznej inteligencji.

6. *MULTIEMO* — NARZĘDZIE DO WIELOJĘZycznej ANALIZY SENTYMENTU

*MultiEmo*³³ jest nowym, modelowym zbiorem danych utworzonym na cele wielojęzycznej analizy sentymentu. *MultiEmo* wywodzi się

³¹ V. Chaudhary, Y. Tang, F. Guzmán, H. Schwenk, P. Koehn. *Lowresource corpus filtering using multilingual sentence embeddings*, w: O. Bojar, i in. (red.), *Proceedings of the Fourth Conference on Machine Translation (WMT)*, Association for Computational Linguistics 2019.

³² Ł. Garncarek, R. Powalski, T. Stanisławek, B. Topolski, P. Halama, M. Turski, F. Graliński, *LAMBERT: Layout-aware language modeling for information extraction*, w: J. Lladós, D. Lopresti, S. Uchida (red.), *Document Analysis and Recognition — ICDAR 2021*, Springer International Publishing 2020, s. 1–16.

³³ J. Kocoń, P. Miłkowski, K. Kanclerz, *MultiEmo: Multilingual, Multilevel, Multi-domain Sentiment Analysis Corpus of Consumer Reviews*, w: M. Paszynski, D. Kranzlmüller, V.V. Krzhizhanovskaya, J.J. Dongarra, P.M.A. Sloot (red.), *Computational Science — ICCS 2021. ICCS 2021. Lecture Notes in Computer Science*, t. 12743, Springer, Cham 2021.

PolEmo (v. 1), zawierającego 57 466 zdań w 8216 różnych dokumentach/recenzjach/opisach. Dane *PolEmo* zostały ręcznie anotowane w warstwie sentymentu zarówno na poziomie całego tekstu, jak i poszczególnych zdań. Podczas tworzenia *MultiEmo* dane *PolEmo* zostały automatycznie przetłumaczone na dziesięć języków, w tym na język rosyjski. W wyniku zastosowanej anotacji dane korpusowe mogą być analizowane na poziomie całego tekstu lub jego części – zdań i akapitów.

Korpusy *PolEmo* i *MultiEmo* zawierają zbiory recenzji konsumenckich z zakresu usług medycznych, hotelarskich, obsługi konsumenckiej i obszaru powiązanego z działalnością uniwersytecką. Narzędzie *MultiEmo* jest publicznie dostępne na licencji Creative Commons Attribution 4.0 International³⁴.

7. TAGERML

Bezpośrednio na stronie CLARIN-PL użytkownik uzyskuje dostęp do wielojęzycznego anotatora *TagerML*³⁵ z zaimplementowanymi funkcjami obsługi tekstów rosyjskojęzycznych. Ten tager odwołuje się do *UDPipe*³⁶.

8. INFOREX

Również na stronie CLARIN-PL znajduje się kolejne narzędzie *Info-forex*³⁷, które służy do komputerowego wspomaganie ręcznego anotowania własnych zasobów. Tworzący rosyjskojęzyczne korpusy mogą w procesie znakowania rosyjskich tekstów skorzystać z oferowanej w *Infoforex* funkcji automatycznego przypisywania poszczególnym formom sugerowanych morfosyntaktycznych znaków. W tym celu *Infoforex* komunikuje się z tym samym co *TagerML* narzędziem *UDPipe*.

³⁴ Zasoby do pobrania: J. Kocoń, K. Kanclerz, P. Miłkowski, B. Bojanowski, M. Zaśko-Zielińska, *PolEmo 1.0 + MultiEmo-Test 1.0 Multilingual Sentiment Analysis Dataset for KES2020*, CLARIN-PL digital repository 2020, <http://hdl.handle.net/11321/737> (11.11.2021), J. Kocoń, P. Miłkowski, K. Kanclerz, *MultiEmo: Multilingual, Multilevel, Multidomain Sentiment Analysis Corpus of Consumer Reviews*, CLARIN-PL digital repository 2021, <http://hdl.handle.net/11321/798>, (11.11.2021).

³⁵ <http://ws.clarin-pl.eu/taggerml.shtml> (11.11.2021).

³⁶ <http://lindat.mff.cuni.cz/services/udpipe/> (11.11.2021), por. też M. Straka, J. Straková, *UDPipe*...

³⁷ <https://infoforex.clarin-pl.eu/>.

9. KORPUSY WŁASNE UŻYTKOWNIKÓW PLATFORMY CLARIN-PL

CLARIN-PL wspomaga użytkowników w tworzeniu własnych zasobów. Każdy użytkownik infrastruktury CLARIN po załadowaniu swojego jednojęzycznego korpusu na stronę *Repozytorium dSpace* uzyskuje po wstępnym przetworzeniu danych dostęp do swoich zbiorów w przeglądarce webowej *KonText*. W przypadku wielojęzycznych zbiorów niezbędna jest konsultacja z Centrum Wiedzy PolLinguaTec CLARIN-PL, które udziela aktywnego wsparcia merytorycznego wszystkim użytkownikom w planowaniu badań, wyborze właściwych metod i narzędzi.

Infrastruktura CLARIN-PL jest nieustannie integrowana z wieloma światowymi repozytoriami narzędzi i zasobów typu Linked Open Data (otwartych danych połączonych). Dla badacza oznacza to możliwość natychmiastowego dostępu do danych połączonych z CLARIN-PL jedną wielką siecią bez konieczności przeszukiwania przepastnych zasobów internetowych.

10. INNE PLANOWANE ZADANIA

Obok konstrukcji nowych korpusów (ręcznie znakowanych i zrównoleglonych) będą prowadzone prace nad utrzymaniem już utworzonych zasobów, które obejmą korektę dostrzeżonych błędów oraz nieustanne przystosowywanie parametrów do zmieniających się standardów. Ponadto planowane jest utworzenie konstruktora zapytań – wspólnego dla wszystkich języków. Utworzenie takiego narzędzia ułatwi zadawanie złożonych pytań dla każdego reprezentowanego w zasobach wielojęzycznych CLARIN-PL języka bez konieczności uczenia się i stosowania różnie zdefiniowanych zbiorów tagów poszczególnych języków.

11. PODSUMOWANIE

W artykule opisano już dostępne na stronie CLARIN-PL i aktualnie tworzone zasoby z językiem rosyjskim (tj. korpusy wielojęzyczne i pamięci TMX) oraz narzędzia, które umożliwiają analizę zbiorów rosyjskojęzycznych. Dostęp do korpusów wielojęzycznych CLARIN-PL wymaga uprzedniej rejestracji na stronie. Narzędzia *TagerML* i *MultiEmo* są dostępne bez wymaganej rejestracji. Warto zwrócić uwagę,

że zarejestrowani użytkownicy uzyskują dojsście do znacznie większej liczby zasobów (<https://clarin-pl.eu/index.php/zasoby/>), narzędzi i usług (<https://clarin-pl.eu/index.php/uslugi/>). Mogą też skorzystać z oferty *Repozytorium dSpace* (<https://clarin-pl.eu/dSpace/>) i chmury *Clarín Cloud* (<https://nextcloud.clarin-pl.eu/>). Każdy użytkownik, deponując w repozytorium lub chmurze własne materiały, określa ich widoczność dla innych osób oraz przypisuje im właściwe prawa autorskie. Wszelkie zamieszczone w repozytorium cyfrowym *dSpace* zasoby (np. korpusy badawcze) mogą „za kliknięciem” użytkownika zostać automatycznie przetworzone i być dostępne w przeglądarce webowej *KonText* (dot. zasobów jednojęzycznych). W przypadku korpusów wielojęzycznych zachodzi konieczność kontaktu z Centrum Wiedzy Pol-LinguaTec, które służy natychmiastową pomocą użytkownikom. Na stronach CLARIN-PL są przystępnie opisane wszystkie zasoby, narzędzia i usługi. W *Mediatece* (<https://clarin-pl.eu/index.php/mediateka/>) zamieszczono praktyczne instrukcje obsługi narzędzi, materiały warsztatowe, publikacje i prezentacje. Również tam znajdzie użytkownik linki do wielu projektów z zakresu e-humanistyki w Polsce. Konsorcjum CLARIN-PL publikuje na kanale YouTube filmy instruktażowe oraz nagrania z konferencji i warsztatów (https://www.youtube.com/channel/UCqrhEITxu8_MIWPnFdYomPw/videos).

REFERENCES

- Banasiak, Jakub. “Built-in argument positions in Bulgarian and Polish.” *Cognitive Studies / Études cognitives*, 2021(21), Article 2558, <https://doi.org/10.11649/cs.2558>, 2021.
- Blagoeva, Diana, and Jaskot, Maciej, and Sosnowski, Wojciech. „A lexicographical approach to the contrastive analysis of Bulgarian and Polish phraseology.” *Cognitive Studies / Études cognitives*, 2019(19), Article 1923. <https://doi.org/10.11649/cs.1923>, 2019.
- Chaudhary, Vishrav, and Tang, Yuqing, and Guzmán, Francisco, and Schwenk, Holger, and Koehn, Philipp. “Lowresource corpus filtering using multilingual sentence embeddings.” *Proceedings of the Fourth Conference on Machine Translation (WMT)*. Bojar, Ondřej i in. Eds. Florence: Association for Computational Linguistics, 2019.
- Dimitrova, Ludmila, and Koseska-Toszewa, Violetta, and Roszko, Danuta, and Roszko, Roman. “Bulgarian-Polish-Lithuanian Corpus: Current development.” *International Workshop: Multilingual resources, technologies and evaluation for Central and Eastern European languages held in conjunction with the International Conference RANLP-2009: Proceedings*. Vertan, Cristina, and Piperidis, Stelios, and Paskaleva, Elena, and Slavcheva, Milena. Eds. Bulgaria. Borovets, 2009: 1–8.

- Dimitrova, Ludmila, and Koseska-Toszewa, and Violetta, Roszko, and Danuta, and Roszko, Roman. “Trilingual Aligned Corpus: Current state and new applications.” *Cognitive Studies / Études cognitives* 2014, no. 14: 13–20.
- Duszkin, Maksim, and Roszko, Danuta, and Roszko, Roman. “New parallel corpora of Baltic and Slavic languages — Assumptions of corpus construction.” *Lecture Notes in Artificial Intelligence LNAI 12848: Text, Speech, and Dialogue TSD 2021*. Ekštejn, Kamil, and Pártl, František, and Konopík, Miloslav. Eds. Cham: Springer International Publishing, 2021: 173–183. DOI: https://doi.org/10.1007/978-3-030-83527-9_15.
- Garncarek, Łukasz, and Powalski, Rafał, and Stanisławek, Tomasz, and Topolski, Bartosz, and Halama, Piotr, and Turski, Michał, and Graliński, Filip. “LAMBERT: Layout-aware language modeling for information extraction.” *Document Analysis and Recognition — ICDAR 2021*. Lladós, Josep, and Lopresti, Daniel, and Uchida, Seiichi. Eds. Cham: Springer International Publishing, 2020: 1–16.
- Kisiel, Anna, and Koseska-Toszewa, and Violetta, Kotsyba, and Natalia, Satola-Staškowiak, Joanna, and Sosnowski, Wojciech. *Polish-Bulgarian-Russian Parallel Corpus*. CLARIN-PL digital repository, 2016, <http://hdl.handle.net/11321/308> (11.11.2021).
- Kocoń, Jan, and Miłkowski, Piotr, and Kanclerz, Kamil. “MultiEmo: Multilingual, Multilevel, Multidomain Sentiment Analysis Corpus of Consumer Reviews.” *Computational Science — ICCS 2021. ICCS 2021. Lecture Notes in Computer Science*, vol. 12743, Paszynski, Maciej, and Kranzlmüller, Dieter, and Krzhizhanovskaya, Valeria V., and Dongarra, Jack J., and Sloot, Peter M.A. Eds. Cham: Springer International Publishing, 2021.
- Kocoń, Jan, and Kanclerz, Kamil, and Miłkowski, Piotr. *MultiEmo: Multilingual, Multilevel, Multidomain Sentiment Analysis Corpus of Consumer Reviews*. CLARIN-PL digital repository, 2021, <http://hdl.handle.net/11321/798>, (11.11.2021).
- Kocoń, Jan, and Kanclerz, Kamil, and Miłkowski, Piotr, and Bojanowski, Bartosz, and Zaśko-Zielińska, Monika. *PolEmo 1.0 + MultiEmo-Test 1.0 Multilingual Sentiment Analysis Dataset for KES2020*. CLARIN-PL digital repository, 2020, <http://hdl.handle.net/11321/737> (11.11.2021)
- Koseska, Violetta, and Roszko, Roman. “On semantic annotation in CLARIN-PL parallel corpora.” *Cognitive Studies / Études cognitives* 2015, no. 15: 211–236. <https://doi.org/10.11649/cs.2015.016> (11.11.2021).
- Machálek, Tomáš. “Advanced and flexible corpus query interface.” *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*. European Language Resources Association, 2020: 7003–7008.
- Piasecki, Maciej, and Walentynowicz, Wiktor. “MorphoDiTa-based tagger adapted to the Polish language technology.” *Proceedings of Human Language Technologies as a Challenge for Computer Science and Linguistics*. Poznań: LTC 2017, 2017: 377–381.
- Roszko, Danuta, and Roszko, Roman. “Polsko-litewskie korpusy IS PAN i CLARIN-PL.” *Prace Baltystyczne* vol. 7. *Język. Kultura. Literatura*. Birgiel, Nijola, and Roszko, Danuta (eds.). Warszawa: Uniwersytet Warszawski, 2018: 185–205.
- Roszko, Danuta, and Roszko, Roman. “Korpusy wielojęzyczne wkładem Instytutu Sławistyki Polskiej Akademii Nauk w rozwój infrastruktury CLARIN-PL: Przykłady analizy korpusowej nad wołączem.” *Języki słowiańskie dziś — w kręgu kategorii, struktur i procesów*. Banasiak, Jakub, and Kikiewicz, Aleksander, and Mazurkiewicz-Sułkowska, Julia. Eds. Warszawa–Łódź: Instytut Sławistyki PAN — Wydawnictwo Uniwersytetu Łódzkiego, 2021: 281–313.

- Roszko, Roman. "O nowych ręcznie zrównoległonych i znakowanych dwujęzycznych korpusach równoległych oraz ich zastosowaniach." *Acta Baltico-Slavica* 2021, no. 45, article 2576.
- Roszko, Roman, and Sosnowski, Wojciech, and Duszkin, Maksim, and Roszko, Danuta, and Tymoshuk, Roman. *Polish-Russian Parallel Corpus, CLARIN-PL* digital repository, 2018, <http://hdl.handle.net/11321/534> (11.11.2021).
- Simov, Kiril, and Simov, Alexander, and Osenova, Petya. "An XML architecture for shallow and deep processing." *The Proceedings of the ESSLLI 2004 Workshop on Combining Shallow and Deep Processing for NLP*. ESSLLI, 2004: 51–60.
- Sosnowski, Wojciech, and Satoła-Staśkowiak, Joanna. "A contrastive analysis of feminines in Bulgarian, Polish and Russian." *Cognitive Studies / Études cognitives*, 2019 (19), Article 1922. <https://doi.org/10.11649/cs.1922>, 2019.
- Straka, Milan and Straková, Jana. *UDPipe, LINDAT/CLARIAH-CZ* digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, Prag 2016, <http://hdl.handle.net/11234/1-1702> (11.11.2021).

Artykuł został częściowo wsparty przez CLARIN – Common Language Resources and Technology Infrastructure, nr projektu POIR.04.02.00-00C002/19.



TOMASZ SZUTKOWSKI

 ORCID <http://orcid.org/0000-0003-3490-1707>

Uniwersytet Szczeciński

W POSZUKIWANIU KOLOKACJI I ICH ODPOWIEDNIKÓW PRZEKŁADOWYCH Z WYKORZYSTANIEM NARZĘDZI ELEKTRONICZNYCH I DANYCH KORPUSOWYCH

IN SEARCH FOR COLLOCATIONS AND THEIR TRANSLATIONAL EQUIVALENTS
WITH THE USE OF ELECTRONIC TOOLS AND CORPUS DATA

This paper presents a problem of stabilized word combinations (collocations) extraction and determining their translational equivalents in Polish and Russian languages. Multi-word language units have been the subject of linguistic research in various aspects of language for many years. However, still relatively few linguists and translators use in their work electronic corpora data, that in the 21st century should be considered as an important source of linguistic information, including real and potential translational equivalents. Author discusses the possibilities of using software as one of the methods of collocation extraction from text material and analyses selected units in a parallel corpus.

Przełom XX i XXI wieku naznaczony był gwałtownym rozwojem technologicznym w niemal wszystkich sferach ludzkiej działalności. Nie tylko nauki ścisłe i przyrodnicze zaczęły na wielką skalę wykorzystywać coraz większe możliwości obliczeniowe komputerów w celu gromadzenia i przetwarzania ogromnych ilości danych. Świat elektroniki nie pozostał jednak na peryferiach szeroko pojętych nauk humanistycznych, w tym także językoznawstwa, które wkroczyło w nową fazę rozwoju zwłaszcza pod szyldem lingwistyki komputerowej i korpusowej¹.

Na marginesie niniejszych rozważań pozostawiam kwestię genezy i aktualnego stanu rozwoju tej językoznawczej subdyscypliny².

¹ Zob. K. Church, M. Liberman, *The Future of Computational Linguistics: On Beyond Alchemy*, „Frontiers in Artificial Intelligence” 2021, vol. 4, s. 1–18.

² Zob. M. Świdziński, *Lingwistyka korpusowa w Polsce — źródła, stan, perspektywy*, „LingVaria” 2006, nr 1, s. 23–34.

W jej kontekście pragnę skupić się natomiast na kilku istotnych zagadnieniach, jakimi są: status wielowyrzowych jednostek języka (nazywanych niekiedy kolokacjami), możliwości ich wyodrębniania w konkretnych tekstach oraz wykorzystanie danych korpusowych w celu ustalenia optymalnych odpowiedników przekładowych. Z moich kilkunastoletnich obserwacji wynika, że w środowisku wielu językoznawców – szczególnie slawistów i rusycystów – zastosowanie narzędzi cyfrowych w badaniach nad językiem spotyka się z dość stanowczym oporem. Naturalnie, wciąż kompetencja specjalisty jest niezbędna w badaniach naukowych; bez niej choćby ogromna ilość wyekscerpowanych wartościowych danych pozostanie jedynie nieuporządkowanym i nieopracowanym materiałem. Warto jednak zwrócić uwagę na fakt, że dzięki różnorodnym programom, aplikacjom i korpusom ów materiał można nie tylko dużo szybciej pozyskać, ale przede wszystkim już na wstępnym etapie dokonać jego selekcji, segregacji czy klasyfikacji. Sposób wykorzystania technologii komputerowej w badaniach językoznawczych może mieć szeroką skalę i zróżnicowany charakter. Z pewnością nie we wszystkich tego typu działaniach niezbędna będzie profesjonalna kompetencja informatyczna, jak np. znajomość języków programowania, tworzenia złożonych baz danych, filtrów ekscerpcyjnych itp. W wielu sytuacjach wystarczy nawet podstawowa umiejętność obsługi komputera i zapoznanie się z nieskomplikowanymi narzędziami informatycznymi, aby móc skutecznie pozyskiwać oraz interpretować istotne dla badań naukowych dane. Celem niniejszego artykułu jest właśnie krótka prezentacja zautomatyzowanej metody ekscerpcji wielowyrzowych jednostek języka z dowolnych tekstów oraz ustalania ich odpowiedników na materiale polsko-rosyjskich relacji przekładowych.

Problem statusu wielowyrzowych jednostek języka nie jest w światowej lingwistyce ani nowy, ani tym bardziej zaskakujący. Można powiedzieć, że co najmniej od czasów Charlesa Bally'ego i jego koncepcji stylistyki funkcjonalnej, czyli od początku XX wieku, kwestia ustabilizowanych struktur leksykalnych na stałe wpisała się w dość wyraźnie zarysowany nurt badawczy, który później – dzięki m.in. pracom Wiktora Winogradowa – ukształtował współczesną frazeologię³. W skrajnej postaci przyjęła ona formę idiomatyki, jednak to nie tylko idiomy (np. „nie przebierać w słowach”, „być panem sytuacji”, „słomiana wdowa”, „ręka rękę myje”, „wyjść na światło dzienne”) wy-

³ Zob. M.A. Алексеенко, *Ещё раз о нерешённых проблемах фразеологии*, „Slavica Stetinensia” 1998, nr 8, s. 83–103.

kazują specyficzną spoiłość semantyczno-strukturalną. Między nimi a tzw. swobodnymi połączeniami wyrazowymi należy dostrzec bardzo obszerny i zróżnicowany jakościowo zbiór jednostek synkretycznych, peryferyjnych, częściowo zleksykalizowanych, znajdujących się niejako w pośrednim stadium procesu idiomatyzacji⁴ (np. „salon fryzjerski”, „świeżo malowane”, „szczyt kampanii”, „szczebel wojewódzki”, „wspominać nawiasowo”). Proces ten natomiast jest determinowany szczególnego rodzaju zależnościami lub inaczej — jak to ujął jeszcze w XIX wieku Jan Karłowicz — ustalonymi zwyczajem powinowactwami danej jednostki leksykalnej z innymi wyrazami⁵. Badania owych zależności czy powinowactw ma fundamentalne znaczenie dla badań językoznawczych, na co zwrócił uwagę Andrzej Bogusławski — „praktycznie zadaniem głównym”, „centralną kwestią lingwistyczną”, „pierwszoplanowym problemem badawczym” jest poszukiwanie i wyodrębnianie wielowyrzawowych jednostek języka⁶, ponieważ ich liczba przybiera paradoksalnie milionowe wartości, co w porównaniu z jednowyrzawowymi jednostkami jest wielkością niewspółmiernie większą⁷. Aspekt onomazjologiczny badań nad wielowyrzawowymi jednostkami podkreśla szczególnie Wojciech Chlebda:

Jeżeli nazywanie jest aktem wieńczącym proces poznawania tego właśnie, a nie innego, wycinka, obszaru czy aspektu świata, to, chcąc mieć orientację w strukturze świata i naszych skomplikowanych z nim relacji, musimy w nieskończonym procesie poznawczym ustalać, jakie pojęcia o fragmentach tego świata jakim werbalizatorom odpowiadają. Mówiąc konkretniej: ustalmy granice werbalizatora, kształtowy zasięg i składowe danej formy słownej, a dojdziemy do ustalenia zasady odrębności i granic odpowiadającego mu pojęcia — tego właśnie, a nie innego, pojęcia o tym i tylko o tym wycinku, obszarze czy aspekcie rzeczywistości⁸.

⁴ T. Szutkowski, *Współczesna paremiografia rosyjska. Stan. Problemy. Perspektywy*, Volumina.pl, Szczecin 2015, s. 46.

⁵ J. Karłowicz, *Przyczynki do projektu wielkiego słownika języka polskiego*, „Rozprawy i sprawozdania Wydziału Filologicznego Akademii Umiejętności” 1876, t. IV, s. XXVIII–XXX.

⁶ A. Bogusławski, *Uwagi o pracy nad frazeologią*, w: Z. Saloni (red.), *Studia z polskiej leksykografii współczesnej*, t. 3, Wyd. Filii UW w Białymstoku, Białystok 1989, s. 18–19.

⁷ Skalę i wagę tego zjawiska bardzo dobrze oddaje nowatorskie opracowanie leksykograficzne zespołu pod kierunkiem Profesora Wojciecha Chlebdy: W. Chlebda (red.), *Polsko-rosyjski słownik par przekładowych. Tom zbiorczy Podręcznego idiomatyzkonu polsko-rosyjskiego (z. 1–5)*, Wyd. Uniwersytetu Opolskiego, Opole 2014.

⁸ W. Chlebda, *Reprodukty na warsztacie*, w: W. Chlebda (red.), *Na tropach reproduktów. W poszukiwaniu wielowyrzawowych jednostek języka*, Wyd. Uniwersytetu Opolskiego, Opole 2010, s. 9–10.

Wraz z rozwojem badań nad semantyką, strukturą i funkcją związków wyrazowych wykazujących pewien stopień leksykalizacji pojawiły się różne koncepcje dotyczące ich nazewnictwa. Trudno byłoby w granicach jednego artykułu przedstawić choćby zarys tego problemu. Można jednak wymienić co najmniej kilkanaście terminów funkcjonujących we współczesnym językoznawstwie: „frazologizm”, „frazem”, „kolokacja”, „odtworzalny wielowyrazowiec”/„reprodukt”⁹, „peryfraz”, „analizm”, „konstrukcja opisowa”, „multiwerbizm”, „połączenie konwencjonalne”, „połączenie frazeologiczne”¹⁰ i in. Za każdą z tych propozycji terminologicznych stoją mniej lub bardziej przekonujące racje i stanowiska. Specyficzne rozmycie niektórych z tych pojęć wymaga od badacza niemal każdorazowego zdefiniowania używanych terminów. Taka sytuacja ma miejsce przede wszystkim w przypadku frazeologizmu, który w zależności od tzw. wąskiego lub szerokiego rozumienia frazeologii¹¹ może oznaczać diametralnie różne jakościowo byty językowe. Nie wdając się jednak w szczegółowe analizy, należy zauważyć, że wspólnym mianownikiem jednostek języka kryjących się za wskazanymi terminami jest przede wszystkim cecha odtwarzalności oraz ich co najmniej binarna struktura. I to właśnie konglomerat tych dwóch parametrów optymalnie oddaje termin „kolokacja” (z łac. „collocatio” — ‘ułożenie, układ, umieszczenie, ustawienie’), który zaistniał w środowisku brytyjskich językoznawców w latach pięćdziesiątych XX wieku¹². Badacze ci podkreślali głównie występowanie określonych schematów strukturalnych oraz specyficzne predyspozycje w zakresie łączliwości jednego leksemu z innymi leksemami. Istotę kolokacji trafnie oddaje definicja Tona van der Woudena: „Collocation is a term that refers to the mechanism, or fact, that certain words are regularly found in the company of other words”¹³. Można zatem stwierdzić, że każdy leksem niejako wytwarza szczególnego rodzaju sieć wiązań z innymi jednostkami leksykalnymi, preferuje współwystępowanie jednego elementu lub kilku, a z in-

⁹ W. Chlebda, *Reprodukty na warsztacie...*, s. 10.

¹⁰ E. Białek, *Kolokacja w przekładzie. Studium rosyjsko-polskie*, UMCS, Lublin 2009, s. 8.

¹¹ T. Szutkowski, *Współczesna paremiografia rosyjska...*, s. 37.

¹² Zob. J.R. Firth, *Papers in Linguistics 1934–1951*, Oxford University Press, London 1957. Stan badań w zakresie kolokacji w językoznawstwie rosyjskim, polskim i zachodnioeuropejskim omawia Ewa Białek: E. Białek, *Kolokacja w przekładzie...*, s. 8–14.

¹³ T. van der Wouden, *Negative Contexts. Collocation, Polarity and Multiple Negation*, Routledge, London 1997, s. 5.

nymi nie wchodzi w językową koegzystencję (np. „odnieść sukces”, „ponieść porażkę”, ale nie „*ponieść sukces”, „*odnieść porażkę”). W zjawisku tym uczestniczy oddziaływanie — z jednej strony — uzusu, a z drugiej — normy językowej¹⁴. Nie zawsze jednak ich wektory są zwrócone w tym samym kierunku¹⁵. Kolokacjom towarzyszy również zjawisko wariantywności. Ponadto, nie wszystkie muszą odznaczać się wysoką frekwencją — niektóre z nich to okazjonalizmy, językowe efemerydy.

Ustalenie wspomnianych przez Chlebdę granic jednostek językowych jest w zasadzie prostą procedurą jedynie w odniesieniu do pojedynczych leksemów. O wiele trudniej wyodrębnić jednostki nieciągłe, wieloelementowe czy — używając nomenklatury lingwistyki komputerowej — N-gramowe¹⁶ (dwu i więcej-gramowe: bigramy, trigramy), czyli kolokacji. Zadaniem badacza jest więc ustalenie stopnia ich spoistości semantyczno-strukturalnej, a jednym z jej głównych wskaźników będzie zjawisko reprodukowalności w mowie potwierdzone nie tylko frekwencją, ale także wartością tzw. siły kolokacji. Cel ten można osiągnąć, odwołując się do konkretnych danych tekstowych.

Językoznawca staje zatem w pierwszej kolejności przed wcale niełatwym zadaniem ekscerpcji kolokacji. Może je pozyskać na kilka sposobów. Jednym z nich jest wciąż metoda tradycyjna, „ręczna”, czy też — inaczej mówiąc — fiskkowa. Z pewnością bardzo pracochłonna i czasochłonna, gdyż wymaga uważnej lektury oraz analizy mniej lub bardziej obszernego materiału tekstowego. Przyczyny, które utrudniają ekscerpcję kolokacji (wielowyrazowego reproduktu), wymienia i omawia szczegółowo Chlebda. Są to: 1) nieobecność wyraźnie zakreślonych granic początku i końca jednostki; 2) występowanie w strukturze jednostki elementów „widocznych” i „niewidocznych”; 3) wariantywność (wymienność komponentów, możliwość poszerzania lub skracania struktury wyjściowej); 4) możliwość istnienia jednostki rozrzuconej w tekście; 5) polisemia (szczególnie

¹⁴ Zob. M. Bańko, *Słownik dobrego stylu, czyli wyrazy, które się lubią*, PWN, Warszawa 2006; E. Geller, A. Dąbrówka, *Słownik stylistyczny języka polskiego*, Świat Książki, Warszawa 2007.

¹⁵ Wpisując w wyszukiwarkę www.google.pl niepoprawne warianty „*ponieść sukces”, „*odnieść porażkę”, możemy odnaleźć co najmniej kilkadziesiąt przykładów ich użycia w tekstach.

¹⁶ Zob. D. Jurafsky, J. Martin, *Speech and Language Processing*, Prentice-Hall, New Jersey 2008.

neosemantyzacja jednostki już w języku funkcjonującej)¹⁷. Ten niełatwy jak widać proces można nieco przyspieszyć, wykorzystując odpowiednie narzędzia informatyczne, które jednak absolutnie nie zwalniają badacza z niezbędnej czujności, wykorzystania własnej kompetencji oraz lingwistycznej intuicji. Metoda automatycznej ekscerpacji dostarczy określonego językowego „surowca”, wymagającego mimo wszystko dalszej starannej i rzetelnej obróbki w zależności od ustalonego celu badawczego.

W 2010 roku w jednym z artykułów wspólnie z Jurijem Fiedoruskowem¹⁸ omówiliśmy sposób ekscerpacji związków atrybutywnych z wykorzystaniem wyrażeń regularnych (*regular expressions*)¹⁹. Metoda ta wymaga użycia specjalnych kodów, nazywanych inaczej ankodami (ros. аношкинский код), w postaci dwuliterowych identyfikatorów odpowiadających kombinacji parametrów semantyczno-gramatycznych²⁰. Procedura pozyskania zadanych modeli strukturalnych zakłada cztery etapy: 1) przygotowanie bazy ekscerpcyjnej w postaci tekstu lub tekstów (korpusu), 2) wprowadzenie tekstu do modułu analizatora morfologicznego w celu otrzymania odpowiednich danych w postaci tagów (znaczników) — do każdej jednostki leksykalnej zostaje przyłączony odpowiadający jej znacznik gramatyczny; 3) modyfikacja tekstu umożliwiająca zastosowanie filtrów ekscerpcyjnych składających się z ankodów i wyrażeń regularnych; przykładowy tekst w postaci linearnej:

Беседа главного редактора газеты „Тамбовские известия” А.К. Клименко с одним из лидеров Тамбовского городского отделения партии пенсионеров [...]

przyjmuje postać następującego segmentu (wyraz + ankod):

Беседа га
главного йб

¹⁷ W. Chlebda, *Nieautomatyczne drogi dochodzenia do reproduktów wielowyrazowych*, w: W. Chlebda (red.), *Na tropach reproduktów...*, s. 19.

¹⁸ Ю. Федорущков, Т. Шутковски, *Лексико-грамматическая сочетаемость атрибутивных словосочетаний русского языка в контексте методов компьютерной эксцерпции*, w: R. G. Tirado, L. Sokolova, I. Votyakova (red.), *La lengua y literatura rusas en el espacio educativo internacional: estado actual y perspectivas*, Rubiños, Granada 2010, s. 1564–1569.

¹⁹ Zob. J. Hopcroft, R. Motwani, J. Ullman, *Wprowadzenie do teorii automatów, języków i obliczeń*, przekł. B. Konikowska, PWN, Warszawa 2003.

²⁰ www.aot.ru (10.10.2021).

W POSZUKIWANIU KOLOKACJI...

редактора аб
газеты гб
Тамбовские йт
известия еб;

4) zastosowanie zadanego ekscerptora, np. `.*йб$\n.*абv`, który wyselekcjonuje z danej próbki tekstu związki wyrazowe o strukturze ‘przymiotnik + rzeczownik’ w dopełniaczu liczby pojedynczej (np. „транспортного налога”, „прожиточного минимума”, „русского языка”)²¹.

Strukturę przykładowych filtrów ekscerpcyjnych przedstawia poniższa tabela²²:

Филтр екскерпцыйны	Выекскерпованы звыязек выразовы
<code>^.*аа\$\n.*\тйа</code>	красивый стол
<code>^.*аб\$\n.*\тйб</code>	красивого стола
<code>^.*ав\$\n.*\тйв</code>	красивому столу
<code>^.*аг\$\n.*\тйг</code>	красивый стол
<code>^.*ад\$\n.*\тйд</code>	красивым столом
<code>^.*ае\$\n.*\тйе</code>	красивом столе
<code>^.*аж\$\n.*\тйт</code>	красивые столы
<code>^.*аз\$\n.*\тйу</code>	красивых столов
<code>^.*аи\$\n.*\тйф</code>	красивым столам
<code>^.*ай\$\n.*\тйх</code>	красивые столы
<code>^.*ак\$\n.*\тйц</code>	красивыми столами
<code>^.*ал\$\n.*\тйч</code>	красивых столах

Omówiona pokrótce metoda umożliwia ekscerpowanie związków wyrazowych (w tym także kolokacji) o z góry założonej strukturze bez względu na stopień leksykalizacji ich komponentów składowych, co może być z punktu widzenia językoznawcy istotną wadą, gdyż wymaga skrupulatnej analizy całej listy otrzymanych wielowyrazowych struktur języka (selekcję danych może jednak znacząco usprawnić oprogramowanie statystyczne²³). Jej zaletą natomiast jest z pewnością możliwość szybkiego pozyskania danych do badań nad modela-

²¹ Ю. Федорушков, Т. Шутковски, *Лексико-грамматическая сочетаемость...*, s. 1567–1568.

²² Тамże, s. 1568.

²³ Zob. P. Wierzchoń, *Пять bardzo skuteczных (справдzonych) способов на массовое выделение многовыразовых сегментов подозранных о фразе-матичность (czyli репродуктов)*, w: W. Chlebda (red.), *Na tropach reпродуктов...*, s. 106–107.

mi strukturalnymi grup wyrazowych (werbo-nominalnych, werbo-adwerbialnych, atrybutywnych itd.)²⁴.

Poszukiwanie kolokacji ułatwiają obecnie coraz lepiej rozwinięte narzędzia korpusowe. Dla badacza kluczową kwestią jest możliwość wyekscerpowania związków wyrazowych, które charakteryzują się już pewnym stopniem leksykalizacji, a więc nie kwalifikują się do zbioru tzw. swobodnych połączeń wyrazowych. Narodowy Korpus Języka Polskiego udostępnia użytkownikom specjalnie opracowany moduł ekscerpcyjny, nazywany kolokatorem²⁵. Narzędzie to umożliwia wyszukiwanie na podstawie kryterium leksykalnego ośrodka kolokacji. Po wpisaniu w okno kolokatora leksemu „alarm” otrzymujemy następujące wyniki (poniżej zamieszczam jedynie początkowe rekordy z listy rankingowej, zob. ilustrcję na następnej stronie).

Kolokator Narodowego Korpusu Języka Polskiego podaje bardzo istotne dla badacza dane: liczbę wystąpień ośrodka kolokacji, pasujące współwystąpienia wraz z frekwencją, liczbę kolokacji spełniających zadane kryteria. Lista rankingowa jest jednak porządkowana według współczynnika Chi², dzięki któremu oblicza się wskaźnik siły kolokacji²⁶. Współczynnik ten, w przeciwieństwie do rozkładu częstości, pozwala na wyciągnięcie poprawnych wniosków dotyczących identyfikacji rzeczywistych, a nie pozornych wielowyrazowych jednostek języka²⁷, gdyż informuje o statystycznej istotności korelacji między dwiema określonymi zmiennymi, którymi w tym wypadku są komponenty danej kolokacji. Naturalnie, nie można uzyskanych danych traktować bezkrytycznie. Wyekscerpowane jednostki wymagają i tym razem od badacza analizy oraz weryfikacji. Struktura powyższej przykładowej listy rankingowej daje już pewien obraz łączliwości wybranego leksemu. Mamy zatem następujące związki wyrazowe: „fałszywy alarm”, „alarm powodziowy”, „alarm szalupowy”, „alarm przeciwpowodziowy”, „alarm bombowy”, „alarm pożarowy”, „wszcząć alarm”, „alarm wibracyjny”, „alarm przeciwpowodziowy”, „alarm antyterrorystyczny”, „alarm przeciwlotniczy”,

²⁴ Zob. J. Fiedoruszkow, *Metody automatyzacji ekscerpcji konstrukcji atrybutywnych języka rosyjskiego*, w: W. Chlebda (red.), *Na tropach reproduktów...*, s. 59–85.

²⁵ www.nkjp.uni.lodz.pl (10.10.2021).

²⁶ Zob. E. Babbie, *Badania społeczne w praktyce*, PWN, Warszawa 2007.

²⁷ M. Hebal-Jezińska, *Podstawowe zasady korzystania z korpusów przy badaniu języka*, w: W. Chlebda (red.), *Na tropach korpusów...*, s. 29.

W POSZUKIWANIU KOLOKACJI...

Kryteria ośrodka kolokacji:

Maksymalny odstęp: Zachowaj szyk:

Kryteria kolokatu:
 Część mowy: Kontekst z lewej: Kontekst z prawej:
 Wielkość próbek: Min. współwystąpienie:

Zaawansowane

3.	szalupowy	alarm_szalupowy (2).
4.	przeciwpowodziowy	alarm_przeciwpowodziowy (34).
5.	bombowy	alarm_bombowy (22), bombowy_alarm (2).
6.	pożarowy	alarm_pożarowy (3).
7.	wszcząć	wszcząć_alarm (21), wszczęła_alarm (16), wszcząć_alarm (10), wszczęto_alarm (8), wszczęły_alarm (6), alarm_wszcząć (2), wszczęły_alarm (2), alarm_wszcząć (2), alarm_wszcząć (2), wszczęły_alarm (1), alarm_wszcząć (1), alarm_wszcząć (1), wszczęto_alarm (1), wszczęły_alarm (1).
8.	wibracyjny	alarm_wibracyjny (8).
9.	przeciwpożarowy	alarm_przeciwpożarowy (11), przeciwpożarowa_alarm (1).
10.	antyterrorystyczny	alarm_antyterrorystyczny (11), antyterrorystyczny_alarm (2).
11.	przeciwlotniczy	alarm_przeciwlotniczy (10).
12.	próbny	próbny_alarm (14).
13.	podnieść	podnieśli_alarm (24), podniosła_alarm (16), podniósł_alarm (8), podniesiono_alarm (6), podnieść_alarm (6), podniosły_alarm (6), alarm_podniosł (2), podnieście_alarm (5), alarm_podnieiony (3), alarm_podniosła (2), podnieśliście_alarm (2), alarm_podnień (2), podniosłem_alarm (1), podnieście_alarm (1), podniosło_alarm (1), podniosłam_alarm (1), alarm_podnieście (1), podnieśliśmy_alarm (1), podnieście_alarm (1).
14.	wszczynać	wszczynała_alarm (2), wszczynano_alarm (2), wszczynać_alarm (2), wszczyna_alarm (2), alarm_wszczynany (1), wszczynali_alarm (1), wszczynając_alarm (1).
15.	podnieść	podniosła_alarm (16), podnieśli_alarm (8), alarm_podniosła (2), podniosła_alarm (1), ogłoszono_alarm (2), ogłosil_alarm (10), ogłoszony_alarm (2), ogłosić_alarm (2), alarm_ogłoszono (1), ogłosili_alarm (2), ogłoszą_alarm (2), ogłosi_alarm (2), ogłoszł_alarm (2), ogłosła_alarm (2), ogłosio_alarm (2), ogłosili_alarm (1), ogłosiliśmy_alarm (1), alarm_ogłoszili (1), alarm_ogłoszono (1), zarządźono_alarm (2), zarządzi_alarm (2), zarządź_alarm (2), zarządź_alarm (2), zarządź_alarm (2), zarządź_alarm (2).
16.	zarządzić	zarządził_alarm (1), zarządza_alarm (2), zarządzić_alarm (2), zarządź_alarm (2), zarządź_alarm (2).
17.	spłoszyć	spłoszył_alarm (4), alarm_spłoszył (1).
18.	ogłaszać	ogłaszam_alarm (1), ogłasza_alarm (2), ogłaszamy_alarm (2), ogłaszając_alarm (1), ogłaszającego_alarm (1), alarm_ogłasza (1), alarm_ogłaszając (1).
19.	włączyć	włączył_alarm (12), włączyła_alarm (4), włączył_alarm (4), włączy_alarm (4), alarm_włączył (4), włączony_alarm (2), włącznie_alarm (1), włączyć_alarm (1), włączylimy_alarm (1).
20.	bojowy	alarm_bojowy (11), bojowe_alarm (1).
21.	odwołać	alarm_odwołano (12), alarm_odwołany (7), odwołano_alarm (7), odwołał_alarm (3), odwołać_alarm (2), odwołany_alarm (2), odwołali_alarm (1), odwołalem_alarm (1), odwołaliśmy_alarm (1), alarm_odwoła (1).

próbnym alarmem”, „podnieść alarm”, „wszczynać alarm” itd. Dane korpusowe odzwierciedla m.in. słownik Mirosława Bańki, który dla celów poprawnościowych odnotowuje najczęstsze kolokacje²⁸. Jak wiadać, wiele z nich pokrywa się z wynikami próbnej ekscerpcji:

alarm (stan alarmowy) • fałszywy ~ • ~ bojowy, bombowy, lotniczy • ~ przeciwpożarowy, przeciwpowodziowy • ogłosić ~, podnieść, wszcząć, zarządzić • bić na ~, uderzyć na ~, narobić ~u pot.

alarm (urządzenie) • włączyć ~, wyłączyć

album • ~ fotograficzny, pamiątkowy, rodzinny

aleja • główna ~a, boczna • szeroka ~a • cienista ~a • brzozowa ~a, kasztanowa, lipowa i in. • ~a parkowa, spacerowa • iść ~ą • ~a prowadzi dokądś, wiedzie

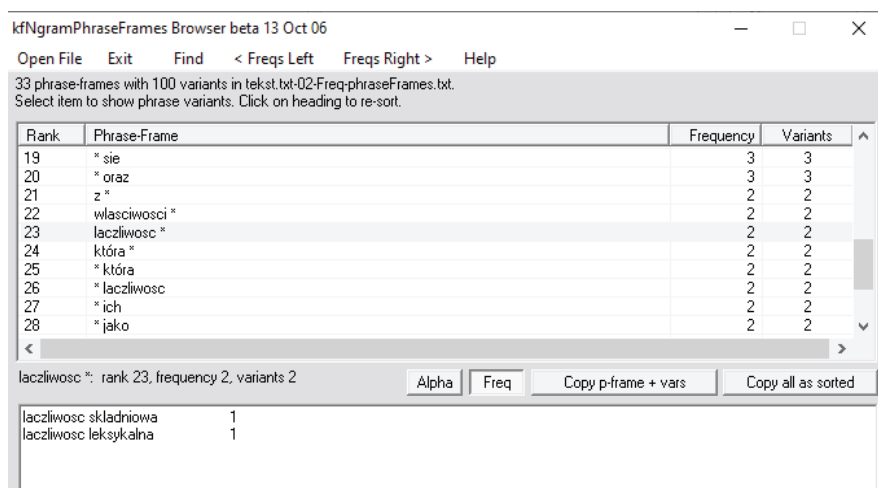
²⁸ M. Bańko, *Słownik dobrego stylu...*, s. 3.

Poza narzędziami korpusowymi językoznawca ma także do dyspozycji inne sposoby pozyskania potrzebnego materiału badawczego, jak na przykład odrębne oprogramowanie, w którym analizie możemy poddać dowolny tekst (w korpusach zasób tekstów jest w danym momencie ograniczony i zamknięty). Mowa o programie kfNgram opracowanym przez Williama Fletchera²⁹ w celu generowania listy n-gramów w plikach tekstowych *txt lub HTML. Wartość „n” przybiera postać dowolnej liczby całkowitej i stanowi określoną sekwencję słów wedle zadanych parametrów. W pierwszym etapie należy sprawdzić, czy plik z przeznaczonym do badań tekstem ma nadany właściwy format (*txt lub HTML). Prosty sposób zmiany formatu jest wklejenie tekstu do znajdującego się w akcesoriach środowiska MS Windows notatnika. Tak sformatowany materiał językowy można poddać analizie. Interfejs programu jest dość prosty w układzie i strukturze. Wybrany plik tekstowy należy wprowadzić do systemu. Przy ustawieniu minimalnego, jednokrotnego wystąpienia, wszystkie składniki tekstu (łącznie ze znakami interpunkcyjnymi) zostaną potraktowane jak oddzielne segmenty. Następnie podzielony na n-gramy plik można poddać analizie frekwencyjnej według ustalonego schematu strukturalnego, a więc np. na bigramy, trigramy, czyli związki wyrazowe 2-, 3-składnikowe. Wyniki analizy są prezentowane w formie listy rankingowej z podaną frekwencją oraz liczbą potencjalnych wariantów (zob. ilustrację na następnej stronie).

Na podstawie analizy fragmentu tekstu o tematyce językoznawczej uzyskujemy bigramy, z których część zakwalifikujemy do zbioru kolokacji, jak np. „łączliwość leksykalna”, „łączliwość składniowa”, „łączliwość wyrazów”, „współwystępowanie wyrazów”, „jednostka wielowyrazowa”, „właściwości składniowe” itd. Należy oczywiście zwrócić szczególną uwagę na to, że program kfNgram podaje wyłącznie współczynnik częstości, który absolutnie nie decyduje o tym, czy dany bigram lub trigram uzyskuje status kolokacji. Na tym etapie znów konieczna jest kompetencja badacza-językoznawcy, który ów status oceni i zweryfikuje (z pewnością kolokacjami nie będą związki: „podobny wyraz”, „jeden język”, „typ wyrazu”). Warto zauważyć, że omawiane narzędzie informatyczne może mieć daleko szersze zastosowanie, znacznie wykraczające poza wyłączone poszukiwanie wielowyrazowych jednostek języka (np. w analizie stylometrycznej).

²⁹ www.kwicfinder.com/kfNgram (12.10.2021).

W POSZUKIWANIU KOLOKACJI...



Rank	Phrase-Frame	Frequency	Variants
19	* sie	3	3
20	* oraz	3	3
21	z *	2	2
22	wlasciwosci *	2	2
23	laczliwosc *	2	2
24	która *	2	2
25	* która	2	2
26	* laczliwosc	2	2
27	* ich	2	2
28	* jako	2	2

laczliwosc *: rank 23, frequency 2, variants 2

laczliwosc skladniowa	1
laczliwosc leksyk.alna	1

Kolokacje w procesie translacji odgrywają fundamentalną rolę. Dotykamy w ten sposób kluczowego zagadnienia, jakim jest pojęcie i istota jednostki tłumaczenia. Białek wyróżnia dwa główne typy tych jednostek:

Przez leksykalną jednostkę przekładu rozumiem pojedyncze leksemy, przez ponadleksykalną zaś — związki wyrazowe, zdanie lub większe fragmenty oryginału, które tylko i wyłącznie jako całość są jednostką sensu, i tylko dla takiej całości może być dobrany odpowiednik w języku przekładu. Przy ponadleksykalnych jednostkach przekładu sens wypływa z integracji sensów wszystkich komponentów, a przekład z różnych względów na poziomie wyrazów nie jest możliwy lub nie jest wskazany³⁰.

W definicji tej badaczka podkreśla wagę integracji sensów komponentów wielowyrazowej jednostki języka (a tym samym jednostki tłumaczenia), której nie można zignorować w procesie przekładu, gdyż w przeciwnym razie takie tłumaczenie może wprowadzić do tekstu w języku docelowym sens niezgodny lub całkowicie sprzeczny z intencją autora oryginału. Praca tłumacza to przede wszystkim ciągle poszukiwanie adekwatnych środków wyrazu sensu, którego nośnikami są jednostki tekstu języka wyjściowego. W kontekście omawianej problematyki kolokacji należy mieć na względzie elementarny fakt — jak podkreśla Chlebda — że mówiący zazwyczaj posługują się w swoich wypowiedziach reprodukowanymi jednostkami ponad-

³⁰ E. Białek, *Kolokacja w przekładzie...*, s. 46.

leksykalnymi. Jednostki jednowyrazowe w zasobach języka tworzą zdecydowanie mniej liczny zbiór³¹. To także ważne spostrzeżenie nakazuje tłumaczowi – z jednej strony – dostrzeganie w tekście oryginału poszczególnych kolokacji jako jednostek będących nośnikiem sensu, z drugiej natomiast – dobieranie do nich ekwiwalentnych jednostek w języku przekładu. Obecnie w epoce intensywnego rozwoju technologii komputerowych sam słownik nie jest już wystarczającym źródłem ekwiwalencji przekładowej³². Korzystanie z korpusów i innych elektronicznych baz danych powoli staje się w codziennej pracy tłumaczy normą i nieodłącznym elementem praktyki zawodowej. Z moich obserwacji wynika, że zwłaszcza adepci niełatwego niewątpliwie rzemiosła translatorskiego, ale także sztuki translatorskiej, niezbyt chętnie korzystają z nowoczesnych narzędzi znacząco ułatwiających i przyspieszających realizację zadania doboru ekwiwalentnych jednostek tłumaczenia (wyjątkiem jest niewątpliwie translator Google).

Bardzo praktycznym narzędziem w ekscerpowaniu kolokacji oraz poszukiwaniu ich tekstowych ekwiwalentów jest czeski korpus elektroniczny InterCorp³³, przydatny nie tylko w eksploracjach języka czeskiego, gdyż dzięki innym podkorpusom umożliwia on analizę jednostek np. w polsko-rosyjskich relacjach przekładowych. Posługując się już wcześniejszym przykładem leksemu „alarm” jako ośrodka kolokacji, można zobrazować zasób potencjalnych wielowyrazowych jednostek języka (w tym przypadku – bigramów) wyekscerpowanych w podkorpusie języka polskiego (zob. ilustrację na następnym stronie)³⁴.

Z tabeli tej badacz może odczytać wiele bardzo wartościowych informacji, ponieważ korpus oblicza jednocześnie kilka istotnych w analizie kolokacji wskaźników statystycznych. Oprócz frekwencji są to: wartość MI3 i MI (prawdopodobieństwo wystąpienia dwóch słów jednocześnie), T-score (tzw. miara kontrastu – im wyższy wy-

³¹ W. Chlebda, *Nieautomatyczne drogi...*, s. 16.

³² O źródłach ekwiwalentów przekładowych pisze Wojciech Chlebda: W. Chlebda, *Korpusologia użytkowa dla początkujących i zaawansowanych*, w: W. Chlebda (red.), *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*, Wyd. Uniwersytetu Opolskiego, Opole 2013, s. 12–13.

³³ <https://www.korpus.cz> (15.10.2021).

³⁴ https://www.korpus.cz/kontext/collx?maincorp=intercorp_v13_pl&view-mode=kwic&pagesize=40&attrs=word&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~OuekoW62Iui0&cattr=word&cfromw=-5&ctow=5&cminfreq=3&cminbgr=1&cbgrfns=3&cbgrfns=m&cbgrfns=t&cbgrfns=d&cbgrfns=p&cbgrfns=r&cbgrfns=s&cbgrfns=l&csortfn=d&collpage=1 (15.10.2021).

	Filtr	word	Freq	MI3	MI	T-score	logDice ▼	MI_log_f	relative freq. [%]	min_sensitivity	log likelihood
1.	p/n	Alarm	74	25.915	13.496	8.602	10.012	58.267	20.274	0.037	1255.402
2.	p/n	fałszywy	78	25.546	12.975	8.831	9.977	56.693	14.130	0.039	1261.687
3.	p/n	Fałszywy	58	26.115	14.399	7.615	9.797	58.711	37.908	0.029	1068.886
4.	p/n	włączył	60	23.968	12.154	7.744	9.490	49.964	8.000	0.030	897.740
5.	p/n	alarm	52	22.796	11.395	7.208	9.110	45.242	4.727	0.026	721.352
6.	p/n	Czerwony	34	21.341	11.166	5.828	8.622	39.699	4.033	0.017	460.304
7.	p/n	włączy	25	22.580	13.292	5.000	8.590	43.307	17.606	0.013	415.674
8.	p/n	pożarowy	22	24.421	15.503	4.690	8.486	48.608	81.481	0.011	456.184
9.	p/n	uruchomił	19	20.728	12.232	4.358	8.139	36.644	8.444	0.010	286.030
10.	p/n	przeciwpożarowy	17	23.016	14.841	4.123	8.109	42.896	51.515	0.009	326.740
11.	p/n	wyłączyć	20	18.952	10.308	4.469	7.829	31.382	2.225	0.010	246.472
12.	p/n	biją	16	19.240	11.240	3.998	7.795	31.844	4.244	0.008	218.132
13.	p/n	czerwony	22	18.744	9.825	4.685	7.743	30.806	1.592	0.011	256.287
14.	p/n	wyłączony	15	18.964	11.150	3.871	7.702	30.915	3.989	0.008	202.596
15.	p/n	fałszywego	14	19.170	11.555	3.740	7.672	31.293	5.283	0.007	197.131
16.	p/n	bije	18	18.389	10.049	4.239	7.642	29.589	1.860	0.009	215.290
17.	p/n	Ogłaszam	13	19.632	12.232	3.605	7.638	32.280	8.442	0.007	195.656
18.	p/n	bojowy	13	19.507	12.106	3.605	7.629	31.949	7.738	0.007	193.298
19.	p/n	Ogłosić	11	21.717	14.798	3.317	7.489	36.772	50.000	0.006	210.470
20.	p/n	włączony	13	17.934	10.533	3.603	7.422	27.796	2.600	0.007	164.260

nik wskaźnika, tym większe prawdopodobieństwo, że dana kombinacja słów jest bardziej ustabilizowana, a więc może stanowić kolokację), logDice (tzw. logarytm Dice'a – pomiar częstotliwości bigramu ze średnią częstotliwością jego składowych), frekwencja względna oraz wskaźnik asocjacji (stosunek częstotliwości bigramu do częstotliwości jednego z jego składników)³⁵. W zamieszczonej powyżej liście rankingowej wybrano jako wskaźnik selekcji parametr logDice, którego obliczenia siły kolokacji nie zależą od wielkości korpusu, co z kolei bardzo ułatwia porównywanie danych między różnymi bazami tekstowymi. Wskaźnik logDice przyjmuje wartość od $-\infty$ do $+14$. Unifikując m.in. warianty graficzne (np. mała i duża litera) i gramatyczne, wysoki wskaźnik siły kolokacji wykazują następujące jednostki: „fałszywy alarm”, „włączyć alarm”, „czerwony alarm”, „alarm pożarowy”, „uruchomić alarm”, „alarm przeciwpożarowy”, „wyłączyć alarm”, „bić na alarm”, „ogłaszać/ogłosić alarm”, „alarm bojowy”, „włączony alarm”. Kolokacje te przyjmują wartość logDice w przedziale (7,422, 9,977). W korpusie można zweryfikować kontekstowo każde wyekscerpowane wystąpienie analizowanej jednostki³⁶:

³⁵ Wszystkie wskaźniki wraz ze wzorami zostały dokładnie omówione pod adresem: https://wiki.korpus.cz/doku.php/pojmy:asociacni_miry?redirect=1#mi-score_a_mi3 (15.10.2021).

³⁶ https://www.korpus.cz/kontext/view?maincorp=intercorp_v13_pl&view-mode=align&pagesize=40&attrs=word&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~asWYgci066KI (15.10.2021).

Wyskrytu: 7 | 1 p.m.: 0,06 (wztażeno k celému korpusu) | ARF: 2,89 | Výsledek je setříděn 1 / 1

Vyběr řádků: základní

InterCorp v13 - Polish

<input type="checkbox"/>	._SUBTITLES	Czerwony alarm .	--
<input type="checkbox"/>	._SUBTITLES	- Czerwony alarm w sercu .	--
<input type="checkbox"/>	._SUBTITLES	Ochrona ... tu jadalnia ... mamy czerwony alarm powtarzam czerwony alarm .	--
<input type="checkbox"/>	._SUBTITLES	Ochrona ... tu jadalnia ... mamy czerwony alarm powtarzam czerwony alarm .	--
<input type="checkbox"/>	._SUBTITLES	- Odwołaj czerwony alarm .	--
<input type="checkbox"/>	._SUBTITLES	Czerwony alarm !	--
<input type="checkbox"/>	._SUBTITLES	Czerwony alarm !	--

4 silnie toksyczny. Kratery wskazują na bombardowania z orbity . - Jak dawno temu ? - Według tempa rozpadu promieniotwórczego, osiemset dziewięćdziesiąt dwa lata temu . Musiały tu żyć miliony . Wygląd . - Wylącam silniki . - Odwołaj czerwony alarm . Przydziel drużyny naprawcze - Tom , upewnij się , że B'Elanna ma wystarczającą pomoc w maszynowni . Chcę mieć z powrotem silniki warp . Tak jest . Nasi przyjaciele są nadal na orbicie . - Krążą jak sępy .

Na podstawie analizy poszczególnych przykładów i kontekstów ich użycia wyłania się kolokacja „czerwony alarm” w znaczeniu ‘stan gotowości, mobilizacja’ (synonim kolokacyjny: „czerwony alert”). Podobne badanie można przeprowadzić w innym języku. Mając na względzie polsko-rosyjskie relacje przekładowe, wykorzystamy w tym celu podkorpus języka rosyjskiego. Po wpisaniu w kolokator leksemu „тревога” uzyskujemy następujące wyniki wyszukiwania:

Filtr	word	Freq	MI3	MI	T-score	logDice	log likelihood	min. sensitivity	MI_log_f	relative freq. [%]	
1.	p / n	вызывает	44	21.075	10.156	6.627	9.484	535.521	0.039	38.661	4.977
2.	p / n	сигнал	38	20.855	10.360	6.160	9.440	473.300	0.034	37.953	5.732
3.	p / n	Ложная	17	22.499	14.324	4.123	8.921	328.827	0.015	41.402	89.474
4.	p / n	вызывают	15	17.834	10.021	3.869	8.394	179.292	0.013	27.783	4.532
5.	p / n	Тревога	12	19.262	12.092	3.463	8.365	179.739	0.011	31.016	19.048
6.	p / n	тревога	13	17.649	10.248	3.603	8.275	159.566	0.012	27.046	5.306
7.	p / n	ложная	11	19.505	12.586	3.316	8.266	173.300	0.010	31.276	26.829
8.	p / n	поводу	32	18.240	8.240	5.638	8.204	303.086	0.013	28.812	1.319
9.	p / n	Сигнал	11	18.233	11.315	3.315	8.196	151.924	0.010	28.116	11.111
10.	p / n	воздушной	9	18.136	11.796	2.999	7.956	130.731	0.008	27.162	15.517
11.	p / n	ложной	9	17.709	11.369	2.999	7.932	125.006	0.008	26.178	11.538
12.	p / n	голосе	10	15.633	8.989	3.156	7.695	104.971	0.009	21.556	2.217
13.	p / n	объявлена	7	17.125	11.511	2.645	7.597	98.678	0.006	23.935	12.727
14.	p / n	чувство	16	15.557	7.557	3.979	7.412	136.139	0.008	21.410	0.821
15.	p / n	тепловой	6	16.917	11.748	2.449	7.393	86.697	0.005	22.860	15.000
16.	p / n	Внимание	8	14.617	8.617	2.821	7.359	79.803	0.007	18.934	1.713
17.	p / n	сигналом	6	15.748	10.578	2.448	7.332	76.431	0.005	20.583	6.667
18.	p / n	охватила	6	15.080	9.910	2.447	7.271	70.725	0.005	19.283	4.196
19.	p / n	поднять	8	14.151	8.151	2.818	7.206	74.613	0.007	17.910	1.240
20.	p / n	Повторяю	6	14.458	9.288	2.446	7.186	65.472	0.005	18.074	2.727

Przyjmując ponownie za kryterium siły kolokacji wartość wskaźnika logDice, można wyodrębnić w analizowanych tekstach następujące kolokacje: „вызывать тревогу”, „сигнал тревоги”, „ложная тревога”, „тревога по поводу”, „воздушная тревога”, „тревога в голосе”, „объявить тревогу”, „чувство тревоги”, „тревога по тепловой защите”, „тревога охватила кого-либо”, „поднять тревогу” i in.

W POSZUKIWANIU KOLOKACJI...

Korpus InterCorp umożliwia również wyszukiwanie ekwiwalentów nie tylko na poziomie jednostek jednowyrazowych, ale także całych kolokacji. Analiza poszczególnych kontekstów daje tłumaczowi szerokie spektrum potencjalnych translatów, znacznie w wielu przypadkach wykraczających poza dane słowników dwujęzycznych. Po wpisaniu w okno wyszukiwarki kolokatora jednostki „czerwony alarm”, korpus wyświetla następujące dane³⁷:

Wyszukiwarka: 7 | l.p.m.: 0,06 (wztażeno k celému korpusu) | ARF: 2,89 | Výsledek je setříděn 1 / 1

Výběr řádků: základní ▾

InterCorp v13 - Polish		InterCorp v13 - Russian	
<input type="checkbox"/>	._SUBTITLES Czerwony alarm .	._SUBTITLES	Пожар .
<input type="checkbox"/>	._SUBTITLES - Czerwony alarm w sercu .	._SUBTITLES	Кардиологам приготовиться .
<input type="checkbox"/>	._SUBTITLES Ochrona ... tu jadalnia ... mamy czerwony alarm ... powtarzam czerwony alarm .	._SUBTITLES	Центральная охрана , это столовая . У нас чрезвычайная ситуация . Повторяю : Чрезвычайная ситуация .
<input type="checkbox"/>	._SUBTITLES Ochrona ... tu jadalnia ... mamy czerwony alarm ... powtarzam czerwony alarm .	._SUBTITLES	Центральная охрана , это столовая . У нас чрезвычайная ситуация . Повторяю : Чрезвычайная ситуация .
<input type="checkbox"/>	._SUBTITLES - Odwołaj czerwony alarm .	._SUBTITLES	Я отработаю за вас остающуюся часть смены .
<input type="checkbox"/>	._SUBTITLES Czerwony alarm !	._SUBTITLES	Тревога !
<input type="checkbox"/>	._SUBTITLES Czerwony alarm !	._SUBTITLES	Тревога !

1 / 1

« не хотела причинить ему вред , но он не послушал меня и не остановился . Эти идиоты видели слишком много старых фильмов о торжме . 28-е марта - 16 дней до поединка Иди к телефону и объяви чрезвычайную ситуацию . Центральная охрана , это столовая . У нас чрезвычайная ситуация . Повторяю : Чрезвычайная ситуация . Айсмен ! Айсмен . До того , как ты попал сюда , мы считали тебя героем . Но потом ты приехал и вел себя так , ▶

Wyłączając z analizy tłumaczenie opisowe lub opuszczenia, na podstawie wyekscerpowanych danych można ustalić co najmniej dwie pary przekładowe: „czerwony alarm” → „чрезвычайная ситуация”, „czerwony alarm” → „тревога”. W podkorpusie języka rosyjskiego możemy co prawda odnaleźć jednostkę „красная тревога”, jednak w korpusie paralelnym nie ma ani jednego przykładu polskiego translatu:

Wyszukiwarka: 5 | l.p.m.: 0,19 (wztażeno k celému korpusu) | ARF: 2,45 | Výsledek je setříděn 1 / 1

Výběr řádků: základní ▾

<input type="checkbox"/>	._SUBTITLES пожар . - Пожар ! Ты сукин . Красная тревога . Повторяю - красная тревога . - Пожар ! -
<input type="checkbox"/>	._SUBTITLES красная тревога . - Пожар ! - Пожар ! Красная тревога ! Готовы ? Бам ! Торопитесь ! Торопитесь ! Пошли
<input type="checkbox"/>	._SUBTITLES она еще перенесет . Это еще что такое ? Красная тревога ! Красная тревога ! Красная тревога ! Похуже , на заводе произошла авария
<input type="checkbox"/>	._SUBTITLES Это еще что такое ? Красная тревога ! Красная тревога ! Похуже , на заводе произошла авария . Назад !
<input type="checkbox"/>	._SUBTITLES покрова планеты ? - 78.6 % Всем станциями . Красная тревога . Переходим на планетарные запасы энергии . Минстер президент ,

1 / 1

³⁷ https://www.korpus.cz/kontext/view?maincorp=intercorp_v13_pl&viewmode=align&pagesize=40&attrs=word&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~buIQek8Sk6o8 (15.10.2021).

Ekwiwalencję tych dwóch kolokacji należałoby zatem zbadać na podstawie kontekstów ich użycia odrębnie w języku polskim i języku rosyjskim. Warto zwrócić uwagę na to, że słowniki polsko-rosyjskie nie odnotowują tej kolokacji. W takich sytuacjach wartość danych korpusowych jest tym bardziej wymierna i cenna.

Z powyższych rozważań wypływa kilka zasadniczych wniosków. Po pierwsze, w badaniach językoznawczych zdecydowanie większą uwagę należy poświęcić wielowyzrazowym jednostkom języka w różnych aspektach — formalnym, semantycznym i funkcjonalnym (pragmatycznym). W wielu ośrodkach naukowych widać od kilkunastu lat postęp tych badań i ich konkretne rezultaty. Przykładem tego jest przede wszystkim seria idiomatykonu, nad którą intensywnie pracuje opolski zespół filologów pod kierunkiem Profesora Wojciecha Chlebdy. Powyższy postulat nie wynika absolutnie z subiektywnych preferencji lub językoznawczej „mody”, lecz z obiektywnych przesłanek dotyczących ontologii języka naturalnego i jego specyfiki, o czym była mowa na wstępie niniejszego artykułu. Po drugie, mając na względzie fakt, że ilościowo kolokacje przeważają nad jednowyrazowymi jednostkami, proces ich wyodrębniania zwłaszcza dla celów leksykograficznych i przekładowych można znacząco przyspieszyć i ułatwić, stosując odpowiednie narzędzia elektroniczne. Jak starałem się to ukazać, nie są one nazbyt skomplikowane, co powinno — w moim przekonaniu — zachęcić zarówno językoznawców starszego, jak i młodszego pokolenia do korzystania z nich w codziennej pracy naukowej i dydaktycznej. Zaprezentowane dane statystyczne tym bardziej podnoszą jakość analizy, a wnioski wyprowadzone na jej podstawie stają się jeszcze bardziej obiektywne. Po trzecie, elektroniczne bazy danych w postaci korpusów (ale nie tylko) znajdują bardzo praktyczne zastosowanie w tłumaczeniach, a proces ustalania par przekładowych na podstawie realnych danych tekstowych (nie kwestionując absolutnie wartości tradycyjnych słowników dwujęzycznych) nie tylko może przyczynić się do podniesienia jakości tłumaczonych tekstów, ale także jest w stanie realnie wspomóc osiągnięcia współczesnej leksykografii przekładowej. Na koniec warto przywołać coraz bardziej aktualne i popularne w językoznawstwie pojęcie frazeotranslacji, które wydaje się swego rodzaju naukowym spoiwem powyższych konstatacji³⁸. Temu zagadnieniu należałoby jednak poświęcić odrębne opracowanie.

³⁸ Zob. J.-P. Colson, *Computational phraseology and translation studies: from theoretical hypotheses to practical tools*, w: G. Corpas Pastor, Jean-Pierre Colson (red.), *Computational Phraseology*, John Benjamins, Amsterdam-Philadelphia 2020, s. 65–81.

REFERENCES

- Alekseyenko, Mikhail. "Yeshchë raz o nereshënykh problemakh frazeologii." *Slavica Stetinensia*, 1998, no. 8: 83–103 [Алексеевко, Михаил. "Ещё раз о нерешённых проблемах фразеологии." *Slavica Stetinensia* 1998, no. 8: 83–103].
- Babbie, Earl. *Badania spoleczne w praktyce*. Warszawa: PWN, 2007.
- Bańko, Mirosław. *Słownik dobrego stylu, czyli wyrazy, które się lubią*. Warszawa: PWN, 2006.
- Białek, Ewa. *Kolokacja w przekładzie. Studium rosyjsko-polskie*. Lublin: UMCS, 2009.
- Bogusławski, Andrzej. "Uwagi o pracy nad frazeologią." *Studia z polskiej leksykografii współczesnej*. Tom 3. Saloni, Zygmunt. Ed. Białystok: Wyd. Filii UW w Białymstoku, 1989: 13–30.
- Chlebda, Wojciech. "Korpusologia użytkowa dla początkujących i zaawansowanych." *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*. Chlebda, Wojciech, Ed. Opole: Wyd. Uniwersytetu Opolskiego, 2013: 7–15.
- Chlebda, Wojciech. "Nieautomatyczne drogi dochodzenia do reproduktów wielowyrazowych." *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*. Chlebda, Wojciech. Ed. Opole: Wyd. Uniwersytetu Opolskiego, 2010: 15–35.
- Chlebda, Wojciech. Ed. *Polsko-rosyjski słownik par przekładowych. Tom zbiorczy Podręcznego idiomatykonu polsko-rosyjskiego (z. 1–5)*. Opole: Wyd. Uniwersytetu Opolskiego, 2014.
- Chlebda, Wojciech. "Reprodukty na warsztacie." *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*. Chlebda, Wojciech. Ed. Opole: Wyd. Uniwersytetu Opolskiego, 2010: 7–13.
- Church, Kenneth. Liberman, Mark. "The Future of Computational Linguistics: On Beyond Alchemy." *Frontiers in Artificial Intelligence*, 2021, vol. 4: 1–18.
- Colson, Jean-Pierre. "Computational phraseology and translation studies: from theoretical hypotheses to practical tools." *Computational Phraseology*. Corpas Pastor, Gloria. Colson, Jean-Pierre. Eds. Amsterdam-Philadelphia: John Benjamins Publishing Company, 2020: 65–81.
- Fedorushkov, Yuriy, and Shutkovski, Tomash. "Leksiko-grammaticheskaya sochetayemost' atributivnykh slovosochetaniy russkogo yazyka v kontekste metodov komp'yuternoy ekstsersptsii." *La lengua y literatura rusa en el espacio educativo internacional: estado actual y perspectivas*. Tirado, Rafael Guzmán, and Sokolova, Larisa, and Votyakova, Irina. Eds. Granada: Rubiños, 2010. 1564–1569 [Федорушков, Юрий, and Шутковски, Томаш. "Лексико-грамматическая сочетаемость атрибутивных словосочетаний русского языка в контексте методов компьютерной экскерпции." *La lengua y literatura rusa en el espacio educativo internacional: estado actual y perspectivas*. Tirado, Rafael Guzmán, and Sokolova, Larisa Votyakova, and Irina. Eds. Granada: Rubiños, 2010: 1564–1569].
- Fiedoruszkow, Jurij. "Metody automatyzacji ekscerpji konstrukcji atrybutywnych języka rosyjskiego." *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*. Chlebda, Wojciech. Ed. Opole: Wyd. Uniwersytetu Opolskiego, 2010: 59–85.
- Firth, John Rupert. *Papers in Linguistics 1934–1951*, London: Oxford University Press, 1957.

- Geller, Ewa. Dąbrówka, Andrzej. *Słownik stylistyczny języka polskiego*. Warszawa: Świat Książki, 2007.
- Hebal-Jeziarska, Milena. "Podstawowe zasady korzystania z korpusów przy badaniu języka." *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*. Chlebda, Wojciech. Ed. Opole: Wyd. Uniwersytetu Opolskiego, 2013: 17–30.
- Hopcroft, John. Motwani, Rajeev. Ullman, Jeffrey. *Wprowadzenie do teorii automatów, języków i obliczeń*. Warszawa: PWN, 2003.
- Karłowicz, Jan. "Przyczynki do projektu wielkiego słownika języka polskiego." *Rozprawy i sprawozdania Wydziału Filologicznego Akademii Umiejętności 1876*, tom 4: 28–30.
- Jurafsky, Dan. Martin, James. *Speech and Language Processing*. New Jersey: Prentice-Hall, 2008.
- Szutkowski, Tomasz. "Subkompetencja frazeologiczna studentów-neofilologów jako składnik kompetencji językowej tłumacza. Badanie pilotażowe." *Roczniki Humanistyczne*, 2019, t. LXVII, z. 7: 93–109.
- Szutkowski, Tomasz. *Współczesna paremiografia rosyjska. Stan. Problemy. Perspektywy*. Szczecin: Volumina.pl, 2015.
- Świdziński, Marek. "Lingwistyka korpusowa w Polsce — źródła, stan, perspektywy." *LingVaria*, 2006, no. 1: 23–34.
- Wierchoń, Piotr. "Pięć bardzo skutecznych (sprawdzonych) sposobów na masowe wyodrębnianie wielowyrzowych segmentów podejrzanych o frazematyczność (czyli reproduktów)." *Na tropach reproduktów. W poszukiwaniu wielowyrzowych jednostek języka*. Chlebda, Wojciech. Ed. Opole: Wyd. Uniwersytetu Opolskiego, 2010: 87–125.
- Wouden, Ton van der. *Negative Contexts. Collocation, Polarity and Multiple Negation*. London: Routledge, 1997.



ALEKSANDRA ZIĘBA

 ORCID: <https://orcid.org/0000-0001-897-5777>

Uniwersytet Śląski w Katowicach

ПАРАЛЛЕЛЬНЫЙ КОРПУС INTERCORP В ПРОЦЕССЕ УСТАНОВЛЕНИЯ ПОЛЬСКО-РУССКИХ ПАР МЕТАОПЕРАТОРОВ

PARALLEL CORPUS INTERCORP IN THE PROCESS OF ESTABLISHING POLISH-RUSSIAN PAIRS OF METAOPERATORS

The article describes the possibilities offered by the parallel corpus InterCorp in the process of translating Polish metaoperators into Russian and searches in the texts the analogous Russian word constructions. The author raises the problem of the interpretation of such concepts as *reproduct*, *multiword unit* and *phrase* in the publications of Polish linguists who more often emphasize the need to define the above-mentioned units by mono- and multilingual dictionaries. The article refers to the metaoperators proposed by Wojciech Chlebda, which are used as an auxiliary tool in the search process from reproducible texts of multiword units. The possibilities and resources of Polish and Russian collections included in the InterCorp corpus were described, and then selected meta-operators of various categories were analyzed. The article describes the possibilities and resources of Polish and Russian collections included in the InterCorp corpus and then selected metaoperators of various categories were analyzed.

Каждый современный язык насыщен бесконечным разнообразием устойчивых и воспроизводимых выражений, которые каждый из нас автоматически употребляет как в письме, так и в повседневной речи, не задумываясь над тем, откуда мы их заимствовали, и как долго данные фразы мы знаем. Они, как будто, укоренились в нашем сознании, сильно активизируясь именно в тех ситуациях, когда их употребление позволяет выразить тот смысл, который автор хотел передать (например, выражение намерения, проявление эмоций и т.д.). Речь здесь не идёт о всем известных фразеологизмах, которым издавна лингвисты и лексикографы уделали немало внимания, результатом чего является обилие доступных одноязычных и переводных фразеологических словарей. Широкое применение хотя бы разговорных

выражений, типа *Ogarnij się, Daj se siana, Cholera wie* или *Krzyż na drodze* безусловно подтвердит каждый носитель польского языка, аргументируя, что *так просто говорится*. В польском языкознании, и прежде всего, лексикографии данного рода конструкции рассматриваются в рамках понятия *репродукт*. Указанный термин, созданный Войцехом Хлебдой, подразумевает под собой извлеченную из текстов единицу языка, которая характеризуется регулярной повторяемостью в этих текстах в функции своеобразного вербализатора¹. Несмотря на то, что в качестве такой конструкции может выступать и однолексемная единица, польский лингвист сосредотачивается преимущественно на выражениях, состоящих из двух или более элементов. Отсутствие в российском языкознании понятия *репродукт*, дальнейшее использование которого в данной статье может вызывать неоднозначные ассоциации, заставляет нас вместо него употреблять термин *фразама*, который воспринимается синонимично в отношении упомянутого многолексемного образования. При таком подходе фраземы образуют своеобразный набор разнородных единиц, характеризующихся другими структурными и семантическими признаками в отличие от типичных фразеологизмов.

Приступая к работе над подобными конструкциями, необходимо сперва ответить на следующие вопросы: *Где мы можем найти фраземы и при помощи каких критериев следует их выделять? Как объяснить, почему одну многолексемную единицу мы должны воспринимать как фразему, а вторую — нет?*

Фраземы выступают преимущественно в текстах, причём следует учитывать тот факт, что компоненты, влияющие на их полную форму, нередко разделены другими лексемами, вследствие чего многолексемная единица является разброшенной. Извлечение фразем из текстовых ресурсов — это непростой процесс, поскольку, в отличие от однолексемных единиц, фраземы сами по себе графически не выделяются, и нередко требуют сложной формальной и семантической проверки. Не подлежит сомнению то, что именно эта особенность привела к тому, что в словарях отмечена в основном нерепрезентативная часть фразем. Даже

¹ W. Chlebda, *Reprodukty na warsztacie*// W. Chlebda (ред.), *Na tropach reпродуктów. Wposzukiwaniu wielowyrzowych jednostek języka*, Wydawnictwo Uniwersytetu Opolskiego, Opole 2010, с. 9.

самые большие фразеологические словари, как одноязычные, так и переводные, до сих пор фиксируют лишь определённый набор этих сверхлексемных единиц. Если говорить о переводной лексикографии то, как правило, лингвисты рассматривают два типа эквивалентов: словарные (системные) и функциональные (текстовые)². Причиной тому является то, что практически не существует словарей, учитывающих второй вид эквивалентов. Главная трудность заключается в определении отношения эквивалентности между трансландом (единица в исходном языке) и транслатом (переводной эквивалент). Принятые критерии извлечения выражений и принципы лексикографической обработки существенно повлияли на факт, что многие конструкции не могли рассматриваться как фраземы и потому не фиксировались в словарях. Только в последние десятилетия лексикографическая ситуация в сфере многолексемных образований начала несколько улучшаться. Применение других методологических решений и, прежде всего, указание других функций как основных, диаметрально изменило взгляд на понятие фраземы. Признание воспроизводимости и метафоричности в качестве ее фундаментальных признаков повлекло за собой резкий рост количества фразем, что, в свою очередь, подтвердило мнение Анджея Богуславского о том, что многолексемные единицы могут исчисляться миллионами³. Благодаря обращению к воспроизводимости, стало возможным выделение не имеющих признаков типовых фразеологизмов фразем (метафоричность, образность и экспрессивность).

Необходимость отмечать и исследовать фраземы испытывает каждый, кто более или менее профессионально занимается переводом. Несмотря на то, что всё больше лексикографов обращается к проблематике помещения фразем в словарях, по-прежнему сложно найти практические советы по переводу данного типа конструкции, не говоря уже о готовых достоверных переводных эквивалентах. К другим причинам, которые затрудняют возможность быстро и однозначно идентифицировать в тексте фразему, относят способность фразем к замене одних

² A. Charciarek, *Параллельный корпус InterCorp в переводной лексикографии* // „Opera Slavica” 2017, вып. 26, 2, с. 7.

³ A. Bogusławski, *Uwagi o pracy nad frazeologią* // Z. Saloni (ред.), *Studia z polskiej leksykografii współczesnej III*, Wydawnictwo Uniwersytetu w Białymstoku, Białystok 1989, с. 19.

компонентов другими, т.е. новые элементы могут проникать в состав многолексемной единицы, изменяя тем самым их облик. Таким образом, структура фраземы в определённой степени остаётся открытой — часть выражения может меняться в зависимости от контекста.

Как уже было упомянуто, вербализаторы отдельных понятий имеют многолексемную форму, и, если наша цель заключается в установлении того, к каким заключениям, описывающим определённые срезы действительности, пришли носители языка, необходимо научиться определять границы фраземы. В этом случае обращение к автоматическим процедурам является недопустимым, однако существует немало других решений, которые направляют и поддерживают интуитивную деятельность лингвиста⁴. На начальном этапе работы нужно ознакомиться с методами и техниками извлечения фразем из текста, а также со способами верификации не только их формы, но и смыслового значения. Подтверждение воспроизводимости установленной многолексемной единицы возможно лишь только в том случае, когда лингвист располагает огромным количеством текстов, свидетельствующих о многократном употреблении фраземы, учитывая при этом сохранение ее структуры и семантического значения.

Не будет преувеличением сказать, что немалое количество текстов в некоторых местах содержит своего рода «подсказки», которые могут сигнализировать о присутствии фраземы. В процессе извлечения многолексемных образований чрезвычайно полезными оказываются неавтоматические способы, под которыми следует понимать то, что:

1) Первостепенную роль при выделении фразем играет интуиция лингвиста. Следует учитывать такие способности, как языковая компетентность, умение отделять одни лексемы от других и непосредственный контакт с текстом. В данном случае, это человек считается основным «устройством», анализирующим структуру текста. Компьютер может служить вспомогательным инструментом для верификации обнаруженных лексических единиц.

⁴ P. Wierzchoń, *Pięć bardzo skutecznych (sprawdzonych) sposobów na masowe wyodrębnianie wielowyrazowych segmentów podejrzanych o frazematyczność (czyli reproduktów)*// W. Chlebda (ред.), *Na tropach reproduktów...*, с. 87–125.

2) Каждое извлеченное из текстовых ресурсов выражение не может восприниматься как неизменяемый по своей структуре объект. Лингвист может только подозревать, что замеченный комплекс лексем образует потенциальную фразу.

В монографии *Na tropach reproduktów. W poszukiwaniu wielowrazowych jednostek języka* (2010), которая представляет собой первую попытку описания проблематики извлечения фразем из текстовых ресурсов, Войцех Хлебда представляет термин *rotosniki formalne*, под которым понимает следующее:

Przez «pomocniki formalne» rozumiem 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 [...]. Znakiem używanym w tej funkcji jest najczęściej cudzysłów, sformułowaniami zaś — metaoperatory: ogólne i względne⁵.

Наряду с определением «метаоператор» языковед употребляет сокращённую форму данного термина — «оператор». На первый взгляд может показаться, что кавычки представляют собой идеальный и надёжный метод, позволяющий безошибочно обособлять фраземы из текста — между графическими границами кавычек содержится искомая многолексемная единица. Однако, не следует забывать о том, что кавычки — это многофункциональный знак, и его использование и понимание каждый раз требует дополнительной интерпретации. Определяющая функция кавычек — извлекать чужие слова, то есть цитаты, которые обычно заимствованы из чьей-то индивидуальной речи. Поэтому нижеуказанный пример использования кавычек (и того, что находится в границах знака препинания) с точки зрения извлечения многолексемных воспроизводимых образований не будет рассматриваться (примеры из Национального корпуса польского языка):

Papież również zdaje sobie sprawę z nadchodzącej przemiany i uznaje jej konieczność. Andre Frossard, cytując bez cudzysłowu myśl papieża, tak to wyraża: «W chaosie świata widzi on (papież) znaki przyszłego ładu, który nam jeszcze umyka... Ów ład w zrozumieniu i przekonaniu Kościoła może być stworzony jedynie przy uznaniu obecności i potęgi Boga. Ale dążenie do uzyskania lepszego ładu, określane, niejednokrotnie kłamliwie, walką o postęp, towarzyszy człowiekowi od zarania dziejów, różniąc go i w tym od zwierząt» (*Watykan bez tajemnic*)

⁵ W. Chlebda, *Reprodukty na warsztacie...*, s. 21.

В работе над фраземами нас должны интересовать такие выражения, которые отличаются обобщённым характером и присутствуют в сознании большинства носителей определённого языка. В свою очередь, в тексте:

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» («Gazeta Wyborcza»).

– кавычки можно успешно заменить такими операторами, как *jak to się mówi* или *(na) tak zwaną*. Такие конструкции указывают на то, что данная последовательность слов часто используется для обозначения явлений и, следовательно, считается многократной и воспроизводимой. Характеристике операторов и возможностям их обнаружения в тексте и переводе будет посвящена следующая часть статьи, в которой мы постараемся извлечь аналогичные конструкции русского языка.

Как мы уже подчёркивали раньше, зачислить многолексемную единицу в разряд фразем возможно только в том случае, если данная форма и значение выражения присутствуют во многих текстовых ресурсах. Раньше традиционным методом сбора лингвистического материала было ручное создание каталогов в бумажном виде, в которых отмечались одно- или многолексемные единицы в определённых контекстах. Появление национальных корпусов текстов существенно изменило ситуацию и открыло новые возможности в аспекте получения, сортировки и анализа языкового материала. Автоматизация работы лексикографа в сочетании с большими коллекциями текстов повысила не только скорость проведения исследований, но и достоверность полученных данных. Несмотря на то, что решение о концепции лексикографического описания по-прежнему принимает лингвист, статистические данные нередко являются основой для составления конкретных тезисов и выводов. Национальные корпусы постоянно расширяют свои рамки, а в их составе образуются новые подкорпусы, что значительно повышает их репрезентативность, благодаря фиксации текстов из самых разных областей (например, художественные, публицистические, а также записи устной речи). Под понятием *национальный корпус* следует понимать представленное в электронной форме собрание текстов, снабженное научным аппаратом,

т.е. аннотацией⁶. Материалы корпуса, а также дополнительные инструменты, которые можно использовать при анализе языка, в значительной степени относятся к лексическим единицам, что делает данные корпуса наиболее полезными для разработки словарей, тезаурусов, глоссариев или в процессе преподавания иностранного языка⁷.

Для достижения целей, поставленных в настоящей статье, где мы обратимся к методам получения русских операторов и проведения дальнейших исследований, наше внимание главным образом будет сосредоточено на явлении параллельных корпусов (*korpus równoległy*). Параллельный корпус — это собрание текстов и соответствующих им переводов на один (или несколько) иностранных языков. Таким образом, такой корпус состоит из двух, первый из которых включает в себя оригинальные тексты на исходном языке, а второй является его отражением на другом языке. Предполагается, что направление перевода текстов, содержащихся в корпусе, не обязательно должно быть одинаковым для каждого из них. Допускается, чтобы в набор текстов входили некоторые оригинальные тексты на родном языке, а остальные представляют собой переводы с другого языка. Параллельные корпуса в лингвистике, в первую очередь, подходят людям, занимающимся сопоставительными и переводческими исследованиями. Наиболее распространённым типом параллельного корпуса является двуязычный корпус, однако существуют параллельные корпуса, собранные для более чем одного языка. Примером такой текстовой базы является один из лучших славянских корпусов, в состав которого входят переводы для сорока языков. Именно этот корпус и будет служить основным инструментом при анализе польских и русских операторов в дальнейшей части работы.

Параллельный корпус InterCorp включает в себя значительную часть Национального корпуса чешского языка (*Český národní korpus*), т.е. старейшего корпуса славянского языка, который создан в 1994 году. Это, прежде всего, многоязычная база текстов, постоянно обогащаемая новыми текстами (так называе-

⁶ В.А. Плунгян, *Зачем нужен Национальный Корпус русского языка? Неформальное введение // Национальный корпус русского языка: 2003–2005*, с. 6–20.

⁷ В. Lewandowska-Tomaszczyk, *Podstawy językoznawstwa korpusowego*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2005, с. 16.

мый открытый корпус). Корпус InterCorp состоит из двух частей — ядра (польск. *trzon*) и коллекций (польск. *kolekcje*)⁸.

Ядро составляют, прежде всего, художественные тексты, а помимо этого, в корпус также входят автоматически обрабатываемые тексты: вышеупомянутые коллекции. Среди них, в новейшей тринадцатой версии корпуса (2020 год), выделяются публицистические тексты и новости с сайтов Project Syndicate и VoxEurop, юридические тексты корпуса Acquis Communautaire, отчёты с заседаний Европейского парламента Eurparl, субтитры к фильмам из платформы OpenSubtitles и переводы Библии⁹. Необходимо помнить, что указанные источники не фиксируют текстов на всех языках, поэтому для каждой отдельной переводной лексикографии (например, польско-русской или русско-польской) InterCorp создаёт лучшие или худшие условия для работы переводчика.

В данном месте следует подчеркнуть, что первоначальное назначение корпуса — это исследование и работа с чешским языком. Также пользователю предоставляются большие текстовые ресурсы английского языка, что повышает вероятность получения достоверных эквивалентов. Ценность данного корпуса с точки зрения надёжности проведения исследований для других переводных лексикографий определяется размерами корпусов (коллекций) отдельных языков. Дифференцированные условия для поиска эквивалентов хорошо видны в случае польского и русского корпусов. Об обстановке данных языков достаточно подробно можем прочитать в статье Анджея Харцярека, посвященной приложению Treq, которое выводит текстовую базу из корпуса InterCorp:

Следует отметить, что польские, чешские и русские тексты, так называемые коллекции, содержащиеся в InterCorp v9, полностью не совпадают. Русский InterCorp не охватывает текстов из корпусов Eurparl, Acquis Communautaire и Presseurop; он содержит художественные тексты (так называемое ядро), публицистические тексты на сайте Project Syndicate, а также субтитры на

⁸ E. Kaczmarska, *Metody ustalania ekwiwalentów czasownikowych wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2019, с. 45.

⁹ E. Kaczmarska, A. Rosen, *Praktyczny przewodnik po korpusie InterCorp // M. Hebał-Jezińska (ред.), Praktyczny przewodnik po korpusach języków słowiańskich*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2014, с. 207.

сайте OpenSubtitles. Польский InterCorp v9 — один из самых больших по величине иноязычных корпусов и содержит все коллекции InterCorp за исключением публицистических текстов на сайте Project Syndicate¹⁰.

Как можно заметить, автор описывает содержание польских и русских коллекций в девятой версии корпуса InterCorp, но нынешняя тринадцатая версия никаких переломных изменений в этой области не принесла. Различия будут ещё более заметными, если вспомнить тот факт, что каждая новая версия InterCorp с одной стороны обогащает польский и русский корпусы большим количеством новых текстов, но с другой, увеличивает непропорциональность между этими языками. К примеру, двенадцатая версия польского корпуса насчитывала свыше 87 млн слов, а русского — примерно 19 млн лексических единиц и выражений. В свою очередь, в актуальной тринадцатой версии содержится 88 млн слов для польского языка и 20 млн для русского. Совершенно разные объёмы польского и русского корпусов и существенные различия в доступности предполагаемых параллельным корпусом коллекций не должны останавливать исследователей перед попытками проведения исследований в рамках данных языков при помощи InterCorp. Особого внимания заслуживает тот факт, что как польский, так и русский корпусы содержат коллекцию OpenSubtitles, которую, с одной стороны, не следует воспринимать как достоверную запись устной (разговорной) речи, но с другой — текстовый материал в определённой степени отражает спонтанные высказывания в неофициальных ситуациях. В данном месте неслучайно затрагивается тема разговорного языка, который долгие десятилетия оставался вне поля лингвистического внимания и уступал место анализу литературного языка.

Основная проблема языковых корпусов заключалась в отсутствии в них устных текстов, потому что экскерпция и транскрипция разговорной речи — это трудоёмкий и дорогостоящий процесс. Следовательно, ресурсы данной разновидности национального языка были и, скорее всего, будут дефицитными по сравнению с ресурсами письменных текстов, разработка которых не является таким сложным процессом. Однако возмож-

¹⁰ A. Charciarek, *Программа для установления переводных эквивалентов Treq: преимущества и недостатки*, «Studia Rossica Posnaniensia» vol. XLV/1, Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza w Poznaniu, Poznań 2020, с. 121.

ности, предлагаемые корпусами лексикографических текстов, непрерывно расширяются. Лучшим примером является именно InterCorp, каждый год обновляемый и пополняемый новыми текстовыми (также устными) коллекциями. Следует также отметить, что разговорный язык обычно отождествлялся с «небрежным» разнообразием национального языка, что привело к его отсутствию во многих лексикографических работах.

Возвращаясь к теме операторов (например, *jak przyjęło się nazywać, jak się mówi, jak drzewiej mawiano* и т.д.), служащих своеобразными навигаторами в поиске многолексемных воспроизводимых единиц, с большой вероятностью можно предположить, что в некоторых контекстах они могут сигнализировать присутствие разговорной лексической единицы. В тексте рядом с конструкциями данного типа можно ожидать, что перед лингвистом откроется возможность извлечь довольно распространённую в устной речи фразему. Однако, цель нашего предстоящего анализа не будет заключаться в обнаружении в текстах сверхлексемных единиц при помощи операторов в корпусе InterCorp. Используя текстовые ресурсы указанного корпуса, мы постараемся не столько перевести предложенные Войцехом Хлебдой польские операторы на русский язык, сколько найти в процессе перевода аналогичные конструкции, которые могут оказаться пригодными при поиске фразем в русскоязычных текстах.

Войцех Хлебда в работе, посвященной неавтоматическим методам обнаруживания многолексемных образований в тексте, разделяет операторы (другое название — *formuły wskaźnikowe*) на два типа: безотносительные (*ogólne*) и относительные (*względne*)¹¹. Первая разновидность операторов должна передавать информацию о том, что строки слов, к которым она относится, являются частыми и повторяющимися «вообще» выражениями при выполнении своей основной функции, т.е. называть объекты и понятия действительности. Они воспроизводятся всеми носителями данного языка. Функцию таких операторов можно назвать «свидетельствующей» или, по определению Анджея Левицкого, «тестимониальной»¹². К таким конструкциям следу-

¹¹ В. Хлебда, *Метаоператоры в функции поисковой системы в обработке ресурсов Рунета для лексикографических целей*, «Rossica Olomucensia» Vol. XLIX, Olomouc 2010, с. 59.

¹² А.М. Lewicki, *Studia z teorii frazeologii*, Oficyna Wydawnictwo Leksem, Łask

ет отнести операторы, такие как: *jak to się mówi, jak to się zwykle mówić, jak to się przyjęło nazywać, tak zwany, przysłowiowy* и т.д. Их перевод на русский язык не должен вызывать больших трудностей, поэтому нашей задачей будет обнаружение возможных вариантов перевода первого из приведённых операторов при помощи параллельного корпуса InterCorp. Тем самым, мы попытаемся установить аналогичные операторы в другом языке, которые при сохранении критерия ситуативности, находят своё применение в текстах и могут помочь лингвистам извлекать многолексемные русские образования. Для этой задачи мы воспользуемся поисковой системой KonText (<https://kontext.korpus.cz>), которая является одним из трёх интерфейсов корпуса InterCorp.

ЯК ТО СИЄ МÓВІ

Распространённое употребление данного выражения подтверждают корпусные данные Национального корпуса польского языка (<https://nkjp.uni.lodz.pl>). PELCRA, поисковая система НКПЯ, фиксирует 595 вхождений и примеров использования оператора в 74 разных текстах. Данная фраза присутствует в речи или письме носителей польского языка, когда они убеждены в том, что фраза (стоящая перед или после оператора) считается общеизвестной и, тем самым, они оправдывают применение конкретных слов в определённой речевой ситуации. Обратимся сейчас к поисковой системе KonText с целью установления потенциальных русских транслятов оператора *jak to się mówi*, не ограничиваясь выбором конкретной коллекции (поиск во всех доступных текстовых ресурсах корпуса). Результаты запроса представляются следующим образом:

jak to się mówi

- jak to się mówi — как говорится (18 вхождений)
- jak to się mówi — что называется (4 вхождения)
- jak to się mówi — как это говорится (3 вхождения)
- jak to się mówi — недаром говорится (1 вхождение)
- jak to się mówi — так называемый (1 вхождение)
- jak to się mówi — как принято говорить (1 вхождение)
- jak to się mówi — как говорят (1 вхождение)
- jak to się mówi — знаешь пословицу...(1 вхождение)

Поисковый инструмент обнаружил 57 русских переводов, из которых только семь репрезентирует коллекцию OpenSubtitles (остальные принадлежат к художественным текстам — ядро корпуса). Мы решили представить и извлечь из документов только те русские переводы, которые, по-нашему мнению, в большей или меньшей степени могут послужить потенциальным оператором при поиске сверхлексемных образований (во многих примерах конструкция *jak to się mówi* не переводилась). Небольшое количество полученных программой примеров, с одной стороны не может служить основой для проведения достоверных исследований и представления надёжных выводов, но с другой — частотность употребления каждой из обнаруженных лексических единиц можно дополнительно проверить в других источниках. Первый из приведённых транслятов, т.е. *как говорится* кажется наиболее адекватным и дословным переводом анализируемого польского оператора. При помощи Национального корпуса русского языка проверим присутствие каждого извлеченного переводного эквивалента в доступных текстовых ресурсах:

Русский оператор	НКРЯ (количество вхождений)
<i>как говорится</i>	6 680
<i>что называется</i>	4 646
<i>как это говорится</i>	108
<i>недаром говорится</i>	71
<i>так называемый</i>	3 588
<i>как принято говорить</i>	201
<i>как говорят</i>	5 118
<i>знаешь пословицу...</i>	36

Оператор *как говорится* по-прежнему остаётся конструкцией с наивысшей частотностью употребления в текстах, а представленные статистические данные подтверждают использование данных выражений в речи русскоговорящих. Компоненты *говорится* и *называется* (или их варианты) представляют собой основные составляющие операторов, причём последний пример скорее предшествует применению пословицы, которая в рамках воспроизводимых многолексемных образований рассматриваться не будет. Таким образом, перевод одного польского оператора открыл доступ к небольшому репертуару потенциальных русских операторов, употребление и обнаружение которых

в тексте каждый раз требует дополнительной интерпретации и верификации значения во избежание неправильных выводов.

В свою очередь, относительные операторы (*operatory względne*) обозначают то, что воспроизводимость последовательности слов оценивается не «вообще», а по отношению к тем или иным «точкам отсчёта», среди которых Войцех Хлебда выделяет три главных оператора: места, времени и сообщества.

Операторы места указывают, какие «обычаи именованья» преобладали, или преобладают в конкретной стране, местности, территории и регионе. К таким конструкциям следует отнести выражения, типа *jak to mówią na Śląsku, jak się mawiało na Kresach, jak to się zwykło nazywać w stolicy, jak to się mawia w Rosji, jak mawiano w moich okolicach, jak tu się mówi, jak się u nas mówi, było takie lwowskie powiedzenie* и их различные варианты. Для подавляющего большинства операторов данной категории можно сконструировать следующую схему их построения: *jak* + глагол *mówić/mawiać/nazywać* (в прошедшем или настоящем времени) + предлог *w/na*. Употребление выбранной конструкции может указывать на место жительства говорящего. К примеру, если человек применяет в своей речи оператор *jak to się u nas mówi*, то сигнализирует собеседнику о своём происхождении и демонстрирует использование типичных для жителей данного региона выражений. В свою очередь, употребление оператора *jak to mówią na Śląsku* с одной стороны может подчёркивать дистанцию говорящего, т.е. «они так говорят, не мы», а с другой — помогать жителю Силезии представлять существующие в его регионе устойчивые выражения. Это указывает также на полученные ранее знания о языке, которые человек использует во время ситуации общения. Предварительные попытки установления потенциальных русских транслятов для вышеуказанных операторов в корпусе InterCorp оказались unsuccessfulными, поэтому нами было принято решение найти перевод только для следующих компонентов (конструкций):

- *Jak to mówią w...*,
- *Jak mówią w...*,
- *Jak mówią na...*

Количество зафиксированных корпусом вхождений составляет 5 примеров, что не позволяет составить надёжного списка

русских операторов места. Среди доступных переводов нам удалось извлечь следующие конструкции:

Jak to mówią w... — как было принято говорить в...

Jak mówią w... — как говорится в..., как говорят в..., как говорят у нас

Jak mówią na... — как здесь говорят

Как можно заметить, каждый из приведённых транслатов содержит компоненты «как» и определённую форму глагола «говорить» (некоторые вхождения включают в себя и дополнительные лексемы). Можно предположить, что более обширный репертуар польских операторов позволил бы обнаружить другие русские конструкции, использующих в переводе не только готовые польские трансланды, но и вариации составляющих их элементов. Сопоставим частотность употребления полученных переводных эквивалентов в НКРЯ:

Русский оператор	НКРЯ (количество вхождений)
<i>как было принято говорить в...</i>	3*
<i>как говорится в...</i>	624*
<i>как говорят в...</i>	517*
<i>как говорят у нас</i>	63
<i>как здесь говорят</i>	80

Последние две конструкции действительно указывали в текстах на выражения, употребляемые на определённой территории, хотя количество вхождений не является столь же репрезентативным, как ожидалось. В свою очередь, интересными оказались результаты поиска остальных многолексемных образований, для анализа которых были проверены каждый текстовый пример и контекст, в которых данные конструкции использовались. Выражение *как было принято говорить в...* относилось в корпусных текстах не к конкретному району, а определённому отрезку времени, что позволяет зачислить данную конструкцию в разряд темпоральных операторов, чему подтверждением служит следующий фрагмент: «Есть только одно обстоятельство, которое характеризует, **как было принято говорить в советские годы**, моральный облик противостоящих сторон». Несмотря на то, что формулировка *как говорится в...* отличается наивысшей частотностью употребления в текстах, мы зафиксировали только один пример её применения в контексте территории: «Выгнать из училища,

сделать сюргун, **как говорится в Турции**, то есть изгнать из страны; заморить даже в тюрьме, все это можно было сделать с учителем, своим подданным».

Остальные примеры, в большинстве своём, указывали на место записи определённых сентенций, представляли собой безотносительные операторы или ссылались на анекдоты.

И тогда, **как говорится в Евангелии**, восплачутся все племена земные (Мф. 24: 30), каждый человек станет перед лицом этого последнего Суда.

Как говорится в народе, на то и щука в реке, чтобы карась не дремал.

Но, **как говорится в анекдоте**, неприятный осадок остался.

На основании полученных данных можно предположить, что данная конструкция крайне редко употребляется для обозначения места, а данные операторы по своему значению скорее соответствуют таким польским конструкциям, как *jak zostało powiedziane, jak mówią ludzie, jak to było w żarcie*. Таким образом, как следует из вышесказанного, мы не получили подходящих русских аналогов операторов места, но более тщательный анализ каждого вхождения позволил извлечь конструкции, которые можно зачислить к операторам другого типа. Похожая ситуация имела место при поиске выражения *как говорят в...*, где в 134 случаях применения данных слов рядом с конструкцией выступало название государства или слово *деревня*:

Но об этом пока рано, это, **как говорят в Израиле**, другой сипур (другой рассказ). Одевается Бурый, **как говорят в деревне**, под партизан.

В доступных текстовых ресурсах НКРЯ фиксируется употребление указанной фразы преимущественно для описания типичных общеизвестных выражений или употребления поговорок, которые используются определённой группой людей.

Как говорят в таких случаях — лиха беда начало

Цель, **как говорят в президентской администрации**, проста — посмотреть, где чиновников избыток, а где дефицит, какие структуры сокращать, какие — расширять

Таким образом, можно сделать вывод, что найденные нами операторы места намного чаще находят своё применение в раз-

ных контекстах, вследствие чего их следует считать многофункциональными (данные конструкции можно отнести к разным операторам). Материалом для проверки частотности употребления послужил исключительно Национальный корпус русского языка, и, возможно, что более расширенный и многофакторный поиск русских конструкций в других электронных источниках (например, Yandex.ru, Google.ru и т.д.) подтвердил бы их применение в контексте местности.

В свою очередь, упомянутые операторы времени (темпоральные) показывают, как было принято называть какой-то отрезок действительности в определённый период: в какую-то эпоху, в какие-то годы или во время конкретного события. Мы имеем ввиду такие конструкции, как *jak drzewiej mawiano, jak to się za komuny nazywało, jak mawiano za okupacji, mówiąc językiem dzisiejszym, jak to się kiedyś mówiło, co określano mianem, jak byśmy dziś powiedzieli* и тому подобное. Параллельный корпус InterCorp при переводе вышеуказанных операторов и их вариантов зафиксировал только один переводной эквивалент, который обнаружился при попытке установления транслята для конструкции *jak mawiano dawniej*:

Twoje teorie to, **jak mawiano dawniej**, spekulacje myślowe.

Это твои соображения, **как прежде говорили**, умственные выкладки.

Очередной раз имеем дело с ситуацией, когда корпус не смог привести как переводы относительных операторов, так и возможные аналогичные конструкции для русского языка, что, вероятнее всего, является следствием недостаточного количества ресурсов русских текстов и диспропорции в доступных для обоих языков коллекциях. В свою очередь, в НКРЯ содержится только семь документов, в которых выступает данная конструкция, что может говорить о низкой употребительности данного выражения в речи русскоговорящих. Необходимо обратиться к другим источникам с целью верификации.

Последними являются операторы сообщества, связывающие соответствующие выражения с той или иной человеческой общиной: социальной средой, профессиональными группами, меньшинствами различных типов, членами обществ и ассоциаций, субкультурами, небольшим кругом лиц, которые создают свой собственный язык (коды) и коммуникативные привычки.

Операторы сообщества, как правило, довольно точно передают информацию о какой именно группе идёт речь. К ним при-мыкают конструкции, типа *jak mówią sprawozdawcy sportowi, mówiąc językiem polityków, jak zwykli opisywać dziennikarze, mówiąc naszym fachowym żargonem* и т.д. С их помощью не составит труда извлечь из текста лексемы и выражения, характерные для профессиолектов, социолектов, жаргона или идиолектов. Мы решили перевести выбранные компоненты представленных операторов и проанализировать найденные в корпусе тексты. Для запроса *jak mówią...* мы получили и извлекли из источников следующие трансляты:

jak mówią narkomani — как именуют её анашисты
 jak mówią więźniowie — как говорят арестанты
 jak mówią chirurdgowie — как говорят хирурги
 jak mówią socjologowie — как это называют социологи
 jak mówią nianie — как говорят няни
 jak mówią dzieci — как говорят дети
 jak mówią ludzie z Bree — как говорят жители Пригорья
 jak mówią marynarze — как говорят моряки
 jak mówią Amerykanie — как выражаются американцы

Как можно заметить, способы перевода польского выражения не являются слишком разнообразными — в списке преобладает конструкция *как говорят*. Иногда присутствуют и другие аналогичные по функции образования (*как именуют...*, *как это называют...*, *как говорят жители...*, *как выражаются...*). В случае двух последних конструкций следует задуматься над тем, в рамках какого из операторов мы должны рассматривать следующие фразы, т.е. оператор места или сообщества. Для следующего выражения, т.е. *mówiąc językiem...*, мы нашли только одно вхождение, которое по сути своей указывает на фразу, употребляемую в языке коммерсантов:

Ostatnie lwy wyginęły by z głodu na długo przedtem, zanim zdołały by upolować ostatnią zdolną do rozrodu parę antylop czy zebr, czyli **mówiąc językiem handlowym**, poławiacze wielorybów zbankrutowali by na długo przedtem, zanim udało by im się wyniszczyć ostatnie ich egzemplarze.

Последние львы издохли бы от голода гораздо раньше, чем убили бы последнюю пару антилоп или зебр, способную к продолжению рода, так же как — в **переводе на наш коммерческий язык** — китобойный флот обанкротился бы задолго до истребления последних китов.

Таким образом, мы получили небольшой список русских операторов сообщества, дальнейшее исследование которых (при наличии других польских конструкции) позволило бы составить более репрезентативный и достоверный список выражений, сигнализирующих о присутствии многолексемной единицы. Стоит ещё вспомнить о одной из разновидностей относительных операторов, так называемых видовых операторах (*operatory gatunkowe*). Это фразы, которые придают соответствующим формулировкам статус пословиц, поговорок, максим и т.д, включая их в разряд языковых единиц, которые нами воспринимаются как общеизвестные, часто повторяемые, и тем самым, воспроизводимые, например, *jak mówi mądrość ludowa, jak głosi znane porzekadło, zgodnie ze starą maksymą*. В ходе нашего анализа нам также удалось обнаружить похожие конструкции типа *znasz posловицу...*, что позволяет предположить, что операторы одной группы могут в определённых контекстах проникать в другие типы операторов и выполнять их функцию.

Подводя итоги, следует отметить, что Параллельный корпус InterCorp лучше подходит для поиска польско-русских пар безотносительных операторов. Небольшое количество содержащихся в русском корпусе текстов и нерепрезентативное количество полученных вхождений вынуждает дополнительно проверять извлеченные конструкции в других источниках с целью верификации не только структуры оператора, но и частотности его употребления. Как показывает исследование, более результативным процессом является перевод отдельной совокупности компонентов оператора (при ручной обработке каждого примера), а не целых конструкций. Актуальность дальнейшего исследования метаоператоров для обоих языков обосновывается необходимостью извлечения и анализа многолексемных образований, которые должны фиксировать актуальные одноязычные и переводные словари. Их основной недостаток заключается в том, что лексема в них трактуется как основная единица языка, без учёта сверхлексемных единиц и ориентации на текст. Поэтому перед современным лексикографом стоит задача выполнить две важные операции: найти фразему в как можно большем количестве текстов и контекстов, а затем деконтекстуализировать многолексемную единицу и поместить её в словарь. Однако, как уже многократно подчёркивалось, проблема часто возникает в самом начале работы — необходимо правильно извлекать многолек-

семные единицы из языка оригинала, метаоператоры же могут служить надёжным инструментом в этой процедуре. Неадекватные пары переводов нередко являются результатом неточного или неполного определения структуры и метафорического значения исходной единицы. Ошибка в определении значения трансланда автоматически приводит к указанию неадекватного эквивалента. Независимо от лингвистической компетентности лексикографа, каждая извлеченная фразема должна быть тщательно проверена в как можно большем количестве источников. Таким образом, при работе с многолексемными образованиями и метаоператорами не нужно ограничиваться только одним корпусом — лингвист должен использовать всю доступную базу текстовых эквивалентов и методов извлечения искомым единиц. Автоматические способы отбора материала, несомненно, облегчают ручную выборку и повышают скорость проведения исследований, но они всегда будут предоставлять материал, языковой статус которого определит решающая инстанция — человек.

REFERENCES

- Bogusławski, Andrzej. “Uwagi o pracy nad frazeologią.” *Studia z polskiej leksykografii współczesnej III*. Saloni, Zygmunt. (Ed.). Białystok: Wydawnictwo Uniwersytetu w Białymstoku, 1989.
- Charciarek, Andrzej. “Parallel’nyj korpus Intercorp v perevodnoy leksikografii.” *Opera Slavica XXVII*, 2017, vol. 27, 2: 5–17 [Харцярек, Анджей. “Параллельный корпус InterCorp в переводной лексикографии.” *Opera Slavica XXVII*, 2017, vo. 27, 2: 5–17].
- Charciarek, Andrzej. “Programma dlya ustanovleniya perevodnykh Ekvivalentov treq: preimushchestva i nedostatki.” *Studia Rossica Posnaniensnia*, 2020, vol. XLV/1: 121–132 [Харцярек, Анджей. “Программа для установления переводных эквивалентов Treq: преимущества и недостатки.” *Studia Rossica Posnaniensnia*, 2020, vol. XLV/1: 121–132].
- Lewicki, Andrzej Maria. *Studia z teorii frazeologii*. Łask: Oficyna Wydawnictwo Leksem, 2013.
- Lewandowska-Tomaszczyk, Barbara. *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2005.
- Kaczmarek, Elżbieta. *Metody ustalania ekwiwalentów czasownikowych wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2019.
- Kaczmarek, Elżbieta, and Rosen, Alexandr. “Praktyczny przewodnik po korpusie InterCorp.” *Praktyczny przewodnik po korpusach języków słowiańskich*. He-

- bal-Jeziarska, Milena. (Ed.). Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2014: 207–231.
- Chlebda, Wojciech. “Reprodukty na warsztacie.” *Na tropach reproduktów. W poszukiwaniu wielowyrzowych jednostek języka*. Chlebda, Wojciech. Ed. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2010: 9–27.
- Plungyan, Vladimir Aleksandrovich. “Zachem nuzhen Natsional’nyj Korpus ruskogo yazyka? Neformal’noye vvedeniye.” *Natsional’nyy korpus russkogo yazyka: 2003–2005*: 6–20 [Плунгян, Владимир Александрович. “Зачем нужен Национальный Корпус русского языка? Неформальное введение.” *Национальный корпус русского языка. 2003–2005*: 6–20.]
- Wierzchoń, Piotr. “Pięć bardzo skutecznych (sprawdzonych) sposobów na masowe wyodrębnianie wielowyrzowych segmentów podejrzanych o frazematyczność (czyli reproduktów).” *Na tropach reproduktów. W poszukiwaniu wielowyrzowych jednostek języka*. Chlebda, Wojciech. Ed. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2010: 87–125.
- Khlebda, Voytsekh. “Metaoperatory v funktsii poiskovoy sistema vobrabotke resursov Runeta dlya leksikograficheskikh tseley.” *Rossica Olomucensia*, 2010, vol. XLIX: 57–66 [Хлебда, Войцех. “Метаоператоры в функции поисковой системы в обработке ресурсов Рунета для лексикографических целей.” *Rossica Olomucensia*, 2010, vol. XLIX].



YEVHEN LIASHCHEVSKYI

 ORCID: <https://orcid.org/0000-0002-4856-1531>

Uniwersytet Jagielloński w Krakowie

INTEREST IN INDIA AND HINDU TEXTS IN RUSSIAN LITERATURE FROM THE END OF THE 18TH CENTURY TO THE MIDDLE DECADES OF THE 20TH CENTURY

INTEREST IN INDIA AND HINDU TEXTS IN RUSSIAN LITERATURE
FROM THE END OF THE 18TH CENTURY TO THE MIDDLE DECADES OF THE 20TH CENTURY

The article examines the reasons for the emergence of interest in ancient Indian religious and philosophical texts in the Russian cultural sphere, as well as the ways in which Indian texts entered Russian literature. The aim of this article is to show how the first Russian translations of the most influential ancient Indian works, such as *Bhagavadgita* or *Shakuntala*, caused the appearance of a large number of works in Russian literature that in one way or another refer to ancient Indian texts. The article also shows how Russian writers referred to ancient Indian texts in their works, using examples of particular writings.

In the second half of the 18th century, the genre of so-called “oriental story” became extremely popular in Russia. This was, for the most part, due to French representatives of the Enlightenment: Voltaire, Denis Diderot, Jean-Jacques Rousseau, Charles de Montesquieu and others. Russian supporters of their ideas started using the Oriental legends and setting as a carrier of the philosophical ideas they professed; it was also a convenient manner of discussing the most touchy social-political issues.

It was Voltaire whose literary oeuvre had the greatest influence on Russian writers¹. His writings — to mention only *The Princess of Babylon* and *The World as it Goes or the Vision of Babouc* — have be-

¹ See К. Фаззех, *Восточные мотивы и образы в русской прозе и драматургии II-й половины XVIII века*, РУДН, Москва 2014.

come a source of inspiration for many pseudo-oriental works created by Russian authors. Moreover, during this period, some translations and imitations of *The Arabian Nights* and other famous Arabic fairy tales appeared in Western literature. Together with the strategy of veiled speech, it sparked a vogue for the Orient in the 18th century. In Russian literature, oriental motifs proved especially convenient for censorship reasons, as the exotic disguise was indispensable for smuggling the criticism of the real social and political system. However, apart from the satirical works, many adventure stories were created.

The stories mentioned above, translated from Western European languages, had nothing to do with India as such but became one of the catalysts for the popularization of fashion for the Orient in general (and India in particular) which appeared in Europe in the mid-18th century and reached Russia in the early 19th century. Translations of the “gems of Indian literature” into European languages eventually led to a direct interest in India. The first sign of this was the 1788 Russian translation of *Bhagavadgita*, rendered as *Bhagavat-Geta or a conversation between Krishna and Arjuna with commentary*. The translator was Alexander Petrov, a close friend of an outstanding historian and writer Nikolai Karamzin. Petrov translated *Bhagavadgita* from English, and his work was undeservedly overlooked by all the Russian literary historians and critics, who only a hundred years later left their first comments on it. The importance of *Bhagavadgita* consists in the fact that the source text — a 700 pages-long philosophical interlude in the epic *Mahabharata* — is considered by most scholars to be a fundamental compendium of Hinduism and Indian philosophy. Hence, Petrov’s translation contributed to the spread of Hinduism in Russia and was one of the first steps in the development of Russian Indology.

Special attention should be paid to the publication that immediately followed Petrov’s book to become the most famous work of the analyzed period translated into Russian from Sanskrit: the play *Shakuntala* by Kalidasa. Its English translation by William Jones came out in 1789 and the German one by Georg Forster — in 1791. The Russian version by Nikolai Karamzin appeared in 1792. His wonderful foreword to *Shakuntala* reads:

The creative spirit does not live in Europe alone; it is a citizen of the universe. Man is everywhere the same; everywhere he has a sensitive heart, and the mirror of his imagination reflects heaven and earth. Everywhere, Nature is his mentoress and the main source of his pleasures. That is what I thought when re-

ading the *Shakuntala* [...] Almost on every page of this drama, I found the highest beauty of poetry, the perfect, inexplicable tenderness resembling the quiet May evening — the purest, inimitable nature and the art itself [...] For me, Kalidasa is as great as Homer. They both got their brush out of Nature’s hands, and they both depicted Nature...²

Karamzin’s foreword was one of the first literary commentaries on Indian classical texts. Comparing Kalidasa to Homer, Karamzin focused on the beauty and significance of Indian literature, which, in his opinion, was not inferior to the greatest examples of European art. *Shakuntala*’s translation was a cultural shock in Russia because instead of showing a stereotypical land of “wonders” and “righteous sages” it revealed a real and surprisingly unfamiliar image of India. To use Nikolai Grintser’s words, *Shakuntala*’s translation “was a milestone in the development of Russian-Indian literary ties.”³ After a while, the translation of Kalidasa’s play reached the Russian poets, though the reception process was hardly smooth. Indian literature has pushed Afanasy Fet and Fyodor Tyutchev to create poetic works called *Sakontala* in 1847 and 1827, respectively.

During the second half of the nineteenth century, the publication of books and articles concerning India continued unabated. The translations of Sanskrit classics that came out at that time were particularly outstanding. They included excerpts from *Mahabharata* and *Ramayana*, plays by Kalidasa and Sudraki, Bhartari’s poetry, excerpts from *Rigveda*, Indian fairy tales and legends from *Panchatantra* and *Hitopadesha*, etc.

As concerns the most significant translations of the ancient Indian texts published in the 18th and 19th centuries, one must mention the translation of the drama *The Rise of the Moon of the Intellect* by Krishna Misri which appeared in a collective volume in Moscow in 1846 under the title *The Feast of Enlightened Thought*. At that time, however, Vissarion Belinsky spoke about it unflatteringly: “Clearly, this drama is a philosophical allegory; even more so, it is a hundred times duller than the boredom itself.”⁴

It is quite possible that the drama had a relatively low degree of artistic value. The work was, nonetheless, the first venture into trans-

² П. Гринцер, “Саконтала” Карамзина, in. *Политрон: к 70-летию В.Н. Топорова*, Рос. акад. наук, Индрик, Москва 1998, pp. 537–566, http://az.lib.ru/k/kalidasa/text_1792_sceny_iz_sacontaly.shtml (23.02.2020).

³ Ibidem.

⁴ В. Белинский, *Полное собрание сочинений*, т. 10, АН СССР, Москва 1956, p. 196.

lation from Sanskrit by the famous linguist and translator Kaetan Kossovich, who is considered to be one of the founders of Russian Sanskritology. Later, in the *Moskovitianin* magazine, Kossovich published his first translations from Sanskrit provided with scholarly commentary: *Supdas and Upesundas. An Episode from Mahabharata* (1844), *Bhagavat Gita* (1854). *The Tale of Vadiyadgar Jimutavagan*, *The story of Somadeva Bhatta* (1847) and the *Tale of Dhruva* (an excerpt from *Bhagavata Purana*) were published separately. In 1849, a translation of Shudraka's *Mrcchakatika* by under the title of *Vasantazena, an ancient Indian drama*, appeared, and in 1854 a translation of *Bhagavadgita* was published. Moreover, under the direction of Kossovich, one created a Sanskrit textbook, *The Legend of a Hunter and a Pair of Pigeons*, and a Sanskrit-Russian dictionary; unfortunately, the latter work was neither completed nor published. By his scholarly and translatorial activity Kossovich not only laid the foundation for Russian Indology but also kindled an interest in the ancient Indian literature, which became a creative impulse for many poets at the turn of the 19th and 20th centuries.

The phenomenon of translated Indian literature was an inspiration for Apollon Maykov to create a small poem titled *Kupal'shchitsy (Melodiya s beregov Ganga)* (1862). Such poems as, among others, *Tri nochi Buddy* and *Tri vstrechi Buddy* (1885) by Semen Nadson had a similar origin.

Thanks to translations, which over time turned into references, allusions and motifs, Indian culture was more and more “impregnating” Russian literature. Robert Stacy, Professor in Russian Literature at Syracuse University, focused the references to Indian culture and everyday life in Russian literature of the 19th and 20th centuries⁵. He provided a detailed list of poets from the “Pushkin era”⁶ whose works contain the fragments somehow referring to India. These were, among others, *Zamechaniya ob Indii* by Denis Davydov (1784–1839), *Tri rozy* by Dmitry Venevitinov (1805–1827), *Tsyganka* by Yevgeny Baratynsky (1800–1844), *Vasantasena* by Dmitry Oznobishin, and *Predsmertnaya ispoved'* and *Molitva parii* by Apollon Grigoryev (1822–1864). The examples illustrate the importance of translations from Sanskrit made several years earlier, since hardly

⁵ R. Stacy, *India in Russian literature*, Motilal Banarsidass, Delhi 1985.

⁶ It should be noticed that Alexander Pushkin also showed great interest in classical Indian literature, especially in the epic poem of *Ramayana*, which he mentioned in his correspondence with the Russian philosopher-Decembrist Petr Chaadayev.

any poet from the above list could deal with the original Indian texts; it was the translation that played the role of catalyst in the crucible of intercultural exchange.

One of the first Russian writers who consciously decided to delve into the ancient Indian writings in their original form was Konstantin Balmont. The scope of the related activities he undertook was unusually wide. In his original literary works, Balmont glorified the Sanskrit classical texts and sometimes even paraphrased their fragments (for instance, his *Indiyskiy motiv*, *Indiyskiy mudrets*, *Pauk*, *Maya*, and many others were taken from Vedanta). However, he was also engaged in a deep study of the Indian and Buddhist culture. His translations of excerpts from the Rigveda and Atharvaveda, the oldest collections of hymns, were published in 1908. The writer also partially translated the *Upanishads* and some famous Sanskrit plays, such as Kalidasa's dramas and *The Life of Buddha* by Ashvaghosha (1916).

Another important figure in the process of incorporation of the elements of Indian culture into Russian literature was Nikolai Gumilev. The poet continually addressed the theme of the Orient throughout his literary career. Since 1910, almost all of his verse collections (*Romanticheskiye tsvety*, *Zhemchuga*, *Ognennyi stolp*, etc.) have been interspersed with some oriental, mainly Buddhist, motifs and symbols. In many Gumilev's works, the idea of transmigration of souls and reincarnation is prominent, echoing the Buddhist teaching of samsara.⁷ Metempsychosis is embodied in his *Zabludivshiy tramvay* (1921) and *Prapamyat'* (1917). The Indian "topoi" are also present in such poems as *Rayskiy sad* (1911), *Severnyy Radzha* (1908), and *Skazka o korolyakh*, as well as in the play for children *Derevo prevrashcheniy* (1918).

A significant contribution to popularizing Indian culture in Russia was brought by Lev Tolstoy, too. In the late 1870s and early 1880s, during the period of spiritual crisis, the writer made an attempt to find an answer to the burning questions about the meaning of life and the destiny of man. Hence, he turned to rich layers of Indian classical narratives and folklore as well as to the philosophical wisdom of *Vedas* and *Upanishads*. Until his death, Tolstoy paid much attention to Indian culture and was supporting its dissemination in Russia.

Tolstoy took a special interest in Buddhism and the ethical side of its philosophy. Wishing to "teach" Buddhism to the masses, he wrote

⁷ Н. Даренская, *Индия в творчестве Н. С. Гумилева*, ЕФИР, Москва 2017.

two related essays. He began work on the first one, called *Buddha*, in 1886, but sooner rather than later decided to cease working on it: he seems to have found it impossible to fully embrace such a vast and complex philosophical doctrine within the boundaries of a single essay. Nevertheless, after a considerable time, in 1905, a more concise version of *Buddha* was published in the collective volume *Krug chteniya*. From the extant sketches of its earlier and later versions, it can be concluded that Tolstoy's primary goal was not so much to present the Buddha's biography, but rather to accessibly explain the nuances of Indian culture, philosophy, and geopolitical situation. The second Tolstoy's essay was written in early 1910. It is his foreword to the article by Pavel Boulanger entitled *Zhizn' i ucheniye Siddarty Gotamy, prozvannogo Buddoy, t. ye. Sovershenneyshim*. Aside from the above attempts to grasp the principles of Buddhist morality, in 1894 Tolstoy wrote a short story called *Karma*, dedicated to the same topic, in which he outlined the essence of one of the basic concepts of Vedic philosophy — the law of justice or reward. Tolstoy's later story, *Eto ty*, is also connected with the Buddhist idea of the common good.

In addition to the works devoted to Buddhism, in such collective volumes as *Kruzhok dlya chteniya*, *Mysli mudretsov* and *Indiyskaya mudrost'*, Tolstoy has published many proverbs, aphorisms and other quotes from the *Vedas* and *Upanishads*, seeking to popularize Indian philosophy.

Aside from being versed in Indian philosophy, during the last years of his life, Tolstoy openly and actively participated in the social and political life of India: he corresponded with many famous Indian thinkers and journalists, among others with Mahatma Gandhi, Baba Premanand Bharati, Brahma Sri Bhagavata Swami and Vijaya Dharma Suri. According to Abhai Maurya's study titled *Indian and Russian Literary Mutuality*,⁸ he criticized the colonial policy.

It should be stressed that Lev Tolstoy's contribution to the popularization of Indian culture in Russia was hardly confined to the desire to expand the existing basic knowledge about the country and disseminate the information obtained by him in the scholarly and literary circles. He also has succeeded in arousing interest in the high moral principles of Indian philosophy among ordinary people. The writer was so profoundly absorbed with the problems of the Indian

⁸ A. Maurya, *Indian and Russian Literary Mutuality*, Atlantis Press, Paris 2015.

subcontinent and the plight of its inhabitants that he tried to help them with his famous letters; thus, he deserved the deepest respect of the Indian people.

Between the 1920s and death of Joseph Stalin in 1953, there were no examples of intercultural communication in the field of fiction that were of interest to me. The majority of works somehow referring to India and its literature that were written in the above period tended to smack of political ideology, which adversely affected the very nature of intercultural interaction — perhaps the best example is the novel *Opasnyĭ beglets* by Emma Vygodskaya (1948). A closer examination of such texts reveals a tenacious process of bringing the monuments of the Indian literature, philosophy and religion to conformity with the doctrine and day agenda of communist propaganda; it also included the very life of the Indians depicted as “exploited people” in need of an immediate socialist revolution.

The situation did not improve before the imposition of the so-called thaw, i.e. Khrushchev’s internal policy of relaxation in the 1950s. Thus, in 1963 a story such as *Razor’s Edge* by Ivan Yefremov could finally be published, where India was described in detail and realistically enough.⁹

As concerns the Russian school of Indology, after its rapid development at the beginning of the 20th century it was almost obliterated in the 1930s. Its rebirth took place in the early 1960s despite various obstacles put by the authorities trying to curb both the independent fiction and scientific research. It was possible owing to the general relaxation within the Soviet regime after Stalin’s death and the efforts of enthusiasts.

For example, in 1960, *Dhammapada* — one of the most important Indian scriptures containing the collected teachings of Buddha in poetic form as well as the main ethical principles of early Buddhism — was translated from the Pali language by Vladimir Toporov. Roughly in the same period, translations of some ancient *Upanishads*, such as *Brihadaranyaka* and *Chandogya*, were prepared by Alexander Syrkin. The importance of these renditions is evidenced by the fact that the fiction writers almost immediately started drawing on them. In the famous story of the *Monday begins on Saturday* by Strugatsky brothers (1964) one of the characters quotes a fragment from Syrkin’s version:

⁹ See Е. Селиверстова, *Индийские мотивы в русской литературе XX века*, СПбГУ, Санкт-Петербург 2017.

- And now — said the voice, — the next one. “Everything is one Me. This is Me, the world’s Me. Unity with ignorance, which comes from the eclipse of light. It disappears with the development of spirituality.”
- And where does this nonsense come from? — I asked. I wasn’t waiting for an answer. I was sure I was asleep.
- A pericope from Upanishads — the voice told, willing to provide me with a one.
- And what are “Upanishads”? — I was no longer sure I was sleeping.
- I don’t know — said the voice.¹⁰

Apart from that, one should mention the translations of *Mahabharata* by Vladimir Kalyanov and Boris Smirnov. Provided with critical commentary, they appeared at various times. The most detailed information about the history of Soviet Indology has been included in the Grigory Bongard-Levin and Alexei Vigasin’s book titled *Image of India. The Research on Ancient Indian Civilization in the USSR*. Also, a valuable study on the issue can be found in Sergei Serebryan’s monograph *Indology in Russia*, dedicated to the oriental studies in the country before and after the October Revolution.

The analysis shows that in the 19th century and the first decades of the 20th century, extraordinary interest in India, treated as a source of inspiration and beauty, became a trend greatly influencing Russian culture.

In contrast to the previous epoch (XI–XV centuries), in the 19th century, an attempt was made to show the real India. However, its image and teachings in most works remained idealized. Many researchers agree that the growing popularity of the Orient and the demand for its peculiar worldview within the Russian culture coincided and was closely connected with the decline of spirituality in the 19th century and early 20th century.

REFERENCES

- Belinskiy, Vissarion. *Polnoye sobraniye sochineniy*. t. 10. Moskva: AN SSSR, 1956 [Белинский, Виссарион. *Полное собрание сочинений*. т. 10, Москва: АН СССР, 1956].
- Bongard-Levin, Grigoriy, and Vigasin, Aleksey. *Obraz Indii: Izucheniye drevneindiyскоy tsivilizatsii v SSSR*. Moskva: Progress, 1984 [Бонгард-Левин, Григорий, and Вигасин, Алексей. *Образ Индии: Изучение древнеиндийской цивилизации в СССР*. Москва: Прогресс 1984].

¹⁰ Аркадий и Борис Стругацкие, *Понедельник начинается в субботу*, Детская литература, Москва 1965, [http://www.serann.ru/text/ponedelnik-nachinaetsya-v-subbotu-8625\(6.03.2020\)](http://www.serann.ru/text/ponedelnik-nachinaetsya-v-subbotu-8625(6.03.2020)).

INTEREST IN INDIA...

- Darenskaya, Natal'ya. *Indiya v tvorchestve N.S. Gumileva*. Moskva: EFIR, 2017. [Даренская, Наталья. *Индия в творчестве Н.С. Гумилева*. Москва: ЕФИР 2017].
- Faezekh, Karimian. *Vostochnyye motivy i obrazy v russkoy proze i dramaturgii II-y poloviny XVIII veka*. Moskva: RUDN, 2014 [Фаэзех, Каримиан. *Восточные мотивы и образы в русской прозе и драматургии II-й половины XVIII века*. Москва: РУДН 2014].
- Grintser, Pavel. "Sakontala Karamzina." *Polytropon: k 70-letiyu V.N. Toporova*. Moskva: Ros. akad. nauk, Indrik, 1998: 537–566 [Гринцер, Павел. "Саконтала Карамзина." *Polytropon: k 70-letiyu V.N. Toporova*. Москва: Рос. акад. наук, Индрик, 1998: 537–566] <http://az.lib.ru/k/kalidasa/text_1792_sceny_iz_saontaly.shtml/rus>.
- Maugya, Abhay. *Indian and Russian Literary Mutuality*. Atlantis Press, Paris 2015, <<https://www.atlantis-press.com/proceedings/icadce-15/23752>>.
- Seliverstova Yelena. *Indiyskiye motivy v russkoy literature XX veka*. Sankt-Peterburg [Селиверстова, Елена. *Индийские мотивы в русской литературе XX века*. Санкт-Петербург 2017].
- Shifman, Aleksandr. *Lev Tolstoy i Vostok*, Moskva 1971. [Шифман, Александр, *Лев Толстой и Восток*. Москва 1971 <<http://greylib.align.ru/1009/aleksandr-shifman-lev-tolstoj-i-vostok-tolstoj-i-indiya-1-chast.html>>].
- Stacy, Robert. *India in Russian literature*. Delhi 1985.
- Strugatskiye, Arkadiy & Boris. *Ponedel'nik nachinaet-sya v subbotu*. Moskva 1965. [Стругацкие, Аркадий и Борис. *Понедельник начинается в субботу*. Москва 1965 <<http://www.serann.ru/text/ponedelnik-nachinaetsya-v-subbotu-8625>>].



GRZEGORZ PRZEBINDA

ORCID: 0000-0001-9581-0214

Uniwersytet Jagielloński

**ТРИ ВСТРЕЧИ С АНДРЕЕМ СИНЯВСКИМ
(ПАРИЖ — ЖЕНЕВА — КРАКОВ, 1987–1993)**

THREE MEETINGS WITH ANDREI SINYAVSKY. PARIS — GENEVA — KRAKOW (1987–1993)

The above memoirs and literary essay talk about my three meetings with Andrei Donatovich Sinyavsky (1927–1997), Russian writer and literary critic, who was widely known in the 1970s–1990s. Originally, Sinyavsky published under his own name in the USSR, but from 1959 his writings appeared illegally in the West under the pen name “Abram Tertz.” Sinyavsky-Tertz remained in exile in France from 1973 until the end of his life. The first part of my text discusses the contacts that the Krakow underground journal “Arka” and Andrzej Drawicz, the great Polish expert on Russia, established with Russian writers in exile living in Paris during the times of the iron curtain (primarily Irina Ilovayska-Alberti, Natalya Gorbanevskaya, Vladimir Maksimov and Andrei Sinyavsky). I reconstruct from memory my first meeting with Sinyavsky and Maria Rozanova, his wife, at Fontenay-aux-Roses near Paris in 1987. During the writer’s lifetime, Rozanova assisted her husband in his editorial activities. In the second part of the essay I talk about my second meeting with the writer in Geneva in 1988, during the international conference on the Millennium of the Baptism of Kievan Rus’ organized by professor Georges Nivat. In the third and final part, I write about my third meeting with Sinyavsky and Rozanova which took place in Krakow in November 1993, two years after the collapse of the USSR. The result of this last meeting was an interview with Sinyavsky and Rozanova prepared by me and Andrzej Romanowski for the weekly “Tygodnik Powszechny” thirty years ago, and published below with my current commentary in the form of footnotes.

Когда 25 марта 1847 года великий русский, а с тех пор навеки эмигрант Александр Герцен (1812–1870) впервые приехал в Париж, он сразу почувствовал себя, словно на своей второй родине:

Итак, я действительно в Париже, не во сне, а наяву: ведь это Вандомская колонна и Rue de la Paix. В Париже — едва ли в этом слове звучало для меня меньше, чем в слове «Москва». Об этой минуте я мечтал с детства¹.

¹ А.И. Герцен, *Собрание сочинений в тридцати томах*, Издательство Академии Наук СССР, Москва 1954–1960, т. 10 (1956): *Былое и думы 1852–1868, часть V*, с. 16. Польская версия этой статьи была опубликована в ежегодном выпуске «Studia Pigioniana», № 4, 2021 г. Представленный здесь перевод сделан Ириной Адельгейм, опытным переводчиком, в т.ч. романов

О своей первой встрече с Парижем Герцен написал только в 1855 году, но уже в июне 1848 года, в девятом из своих *Писем из Франции и Италии*, он упоминает, что въехал в столицу Франции в марте 1847 года так же, «как некогда въезжали в Иерусалим, в Рим»².

Я вспоминаю эти слова Герцена, который близок мне чрезвычайно, поскольку и сам — разумеется, не будем забывать о соблюдении пропорций — зимой 1987 года въезжал в Париж, словно в духовный Иерусалим. Я впервые оказался на берегах Сены, благодаря щедрой помощи поэта Яна Польковского, редактора и издателя краковского подпольного журнала «Арка», и Анджея Дравича (1932–1997), моего Учителя и друга, короля польских русистов. В то время Ян Польковский, не только редактор журнала, но и крупный поэт, по-прежнему верил в возможность другой, то есть некоммунистической, России, а Дравич еще в 1980 году опубликовал в подпольном издательстве книжечку под названием *Другая Россия*, где лаконично изложил «генеалогию русской эмигрантской литературы»³.

Во время военного положения подпольная краковская «Арка», существовавшая с середины 1983 года, то есть ровесница парижского журнала «Зешиты литератке» — уделяла очень много внимания польско-российским связям, разыскивая в русском Париже «друзей-москалей» нашего времени — если воспользоваться цитатой из *Дзядов* Мицкевича (*Часть III. Отрывок*). И вот в 1985 году в одиннадцатом номере «Арки» некто Марек Лешковский опубликовал хвалебную рецензию, озаглавленную *Военное положение на страницах «Русской мысли»*, на только что (1984) из-

Ольги Токарчук. См. K. Jastrzębska, *Irina Adelgejm — rosyjska tłumaczka i literaturoznawczyni // Między Oryginałem a Przekładem*, Księgarnia Akademicka, Kraków 2021, № 2(52), с. 63–78.

² А.И. Герцен, *Собрание сочинений в тридцати томах*, т. 5 (1955): *Письма из Франции и Италии 1847–1852*, с. 141.

³ A. Drawicz, *Inna Rosja*, Towarzystwo Kursów Naukowych. Wykłady, NOWA, Warszawa 1980, с. 1. Другая статья Дравича *Зарубежная Россия* — первоначально опубликованная в пятнадцатом номере варшавского подпольного журнала «Запис» (июль 1980), сразу после *Саги о носорогах* Владимира Максимова и интервью с Андреем Синявским *Я ощущаю глубокую нетерпимость русской эмиграции* (первоначально для парижской газеты «Монд», 7 июля 1979) — увидела свет на русском языке в парижском ежеквартальнике «Континент» (номер 26 за 1986 год). Добавим, что на обложку этого номера была помещена фотография Чеслава Милоша (1911–2004), в связи с получением им Нобелевской премии по литературе.

данную в Париже книгу Натальи Горбаневской (1936–2013), куда вошли тексты о Польше времен «Солидарности», опубликованные с декабря 1981 по декабрь 1982 года на страницах эмигрантского еженедельника «Русская мысль». В заключительной части своей рецензии Лешковский приводит характерную цитату из заявления русских эмигрантов в защиту польской «Солидарности», к которой добавляет, уже от своего имени, слова искренней благодарности всем нашим друзьям-москалям:

Мы, русские, оказавшиеся на Западе, среди них и те, кто в сорок четвертом году «освобождал» Польшу, всем сердцем и душой на стороне польского народа, его рабочего класса, сумевшего бесстрашно и вопреки всему показать миру, как надо бороться за свои права. Мы на стороне «Солидарности»! Мы с гордостью повторяем сегодня клич, родившийся сто пятьдесят лет назад в дни Польского восстания 1830 года: «За нашу и вашу свободу!» Да здравствует свободная, независимая Польша! Да здравствует «Солидарность»! Эту декларацию подписали Виктор Некрасов, Владимир Максимов, Владимир Буковский, Александр Гинзбург, Вадим Делоне, Эдуард Кузнецов и Наталья Горбаневская. Что еще я могу добавить? В прошлом году редакция ежеквартального журнала «Континент» отметила свое десятилетие. В следующем году «Русская мысль» отметит свое 40-летие. Членам редакционных коллективов, коллегам и обозревателям обоих периодических изданий от имени «Арки» передаю теплые слова признания и восхищения. За все усилия, предпринятые во имя вашей — и нашей — свободы»⁴.

Сегодня мы, разумеется, знаем, что этот Марек Лешковский был не кем иным как Леславом Малешкой, разоблаченным в ноябре 2001 года сотрудником коммунистической Службы безопасности, который выполнял свои тайные функции еще с апреля-мая 1976 г. Сегодня в его биографии в онлайн *Энциклопедии «Солидарности»* можно прочитать, что в восьмидесятые годы Малешка занимался как раз журналом «Арка» и «руководящими структурами оппозиции». Кроме того, ему поручали специальные задания: в 1986 и 1988, во время поездок во Францию, он должен был следить за парижскими эмигрантами периода «Солидарности», а также тамошними «курьерами», поддерживавшими связь с оппозицией в стране⁵. Однако сейчас мне хотелось

⁴ M. Leszkowski, *Stan wojenny na lamach „Russkiej Mysli”, „Arka”* 1985, № 11, с. 128. Заявление русских писателей (декабрь 1981 года) было опубликовано в парижском «Континенте», в номере 31 за 1982 год.

⁵ <https://encysol.pl/es/encyklopedia/biogramy/17471,Maleszka-Leslaw.html> (11.01.2021). Характерно, что Ян Польковский также счел цитируемый мною текст Малешки значимым, раз опубликовал его во второй парижской

ТРИ ВСТРЕЧИ С АНДРЕЕМ СИНЯВСКИМ

бы решительно подчеркнуть, что в процитированном выше суждении Лешковско-Малешки — справедливо апологетическом по отношению к столь дружелюбно настроенной по отношению к нам русской эмиграции в Париже, — я лично не усматриваю и тени руки Службы безопасности. Напротив, это мнение было действительно близко значительной части тогдашней редакции «Арки» и, разумеется, автору этих строк. В конце концов, ведь Ян Польковский направил меня в Париж как человека, активно сотрудничавшего с его журналом в том, что касается российских проблем. Я даже внес свою лепту в подготовку нашумевшего тогда тринадцатого номера «Арки» (начало 1986 года), полностью посвященного русско-польским вопросам. Этот номер грозил якобы краковским подпольным издателям такими неприятностями, что редактор Польковский даже не решился указать название журнала и велел напечатать на первой и четвертой страницах обложки фотокопию титульных страниц еженедельника «Русская мысль» за июнь и июль 1984 года (включая информацию о визите Франсуа Миттерана в Москву и фарсе выборов в народные советы ПНР 17 июня 1984 года).



Обложка и оглавление тринадцатого русского номера подпольной «Арки».

подборке текстов «Арки». См: M. Leszkowski, *Stan wojenny na łamach „Russkiej Myśli”* // ARKA. *Wybór 10–12*. Kraków — 1985. Eseiistyka. Krytyka. Literatura, directeur de la publication A. Podgórczyk-Mietkowska, ред. M. Chudecki, W. Sikora, Aktis, Paryż 1986, с. 202–206.

Угроза для журнала была связана с тем, что если критические статьи в свой адрес во «втором круге обращения» коммунистические власти Польской Народной Республики еще могли стерпеть, то в связях поляков с русской эмиграцией в Париже легко было усмотреть попытку «подрыва союзнических отношений». Тем более, что ранее, в марте 1984 года, после выхода всего пяти или шести номеров «Арки», был допущен неожиданный промах, и несколько сотрудников оказались задержаны Службой безопасности. После этого из состава редакции «Арки» вышел как раз Леслав Малешка, но не потому, что Ян Польковский и Рышард Терлецкий заподозрили, будто он является агентом Службы безопасности, а из-за его безудержной болтливости, хорошо известной всему оппозиционному и полуоппозиционному Кракову. Терлецкий, который был в то время одним из столпов редакции, позже, в 1990 году, вспоминал, что анонимность тринадцатого номера была попыткой перехитрить всевидящее око КГБ:

На рубеже 1985 и 1986 годов вышел второй и пока последний монографический выпуск «Арки». Он появился в результате сотрудничества с редакцией российского еженедельника «Русская мысль» и постоянных связей с его главным редактором Ириной Иловайской-Альберти. Более десятка текстов, в том числе Солженицына, Горбаневской и Сахарова, составили подборку переводов, осуществленных под руководством историка и русиста Анджея Новака. Это был самый странный номер «Арки», поскольку он не содержал никакой информации о нашем участии в его создании, обложка представляла собой коллаж обложек российского журнала, не были указаны дата и место издания. Тринадцатый, «секретный», номер «Арки» — свидетельство наших опасений, что после его публикации мы окажемся под колпаком не только у отечественной СБ, но и у советского КГБ⁶.

Анджей Новак — сегодня автор широко известных книг по истории Польши и польско-российским отношениям — прибег во время работы к помощи моих коллег по Институту русской филологии Ягеллонского университета; я также, по просьбе Польковского, читал перед публикацией польские переводы, довольно — вынужден повторить это и сегодня — сырые. Я даже написал к этому изданию небольшое предисловие, которое в конце концов так и не было напечатано, поскольку его сочли слишком «солженицынским», стирающим, как мне сказали, различия между Россией и СССР. Ниже я сам отчасти критикую Ирину Иловайскую-Альберти (1924–2000) за подобную позицию, и по

⁶ R. Terlecki, *Moja historia „Arki”*, „Arka” 1990, № 29, с. 7.

ТРИ ВСТРЕЧИ С АНДРЕЕМ СИНЯВСКИМ

сей день, что касается этой проблемы, считаю наиболее точным суждение Густава Герлинга-Грудзиньского (1919–2000), высказанное в 1987 году на страницах девятнадцатого номера «Арки» в беседе с Богуславом Соником:

[Соник] — Россия — тема, которая проходит через все ваше творчество. На протяжении многих лет вы последовательно утверждали, что нельзя отождествлять российскую автократию с советским коммунизмом. Эта точка зрения имеет своих сторонников, но также вызывает многочисленные критические замечания, как на Западе, так и на Востоке. Каковы ваши аргументы в этом споре? [ГХГ] — Я — за вышеупомянутое различие. Я однако против того, что коротко именуют синдромом Кухажевского, т.е. мнения, будто корни сегодняшнего советского строя уходят в царскую Россию. Это несомненно так, но, на мой взгляд, несправедливо полностью отождествлять эти две системы. Это неверный путь также и с точки зрения политической, поскольку — нравится нам это или нет — придется в конце концов найти общий язык с русскими, и популярная нынче концепция Центральной Европы нас от этой задачи не освободит. Мы не выиграем, объединившись с Венгрией или Чехией. Говоря о различиях между царской и советской Россией, следует помнить, что в первой существовала настоящая судебная система с настоящими адвокатами, и обвиняемые могли себя защитить. Как это происходит сейчас, всем нам известно. Другой пример: Герцен жил в Лондоне на средства от своих имений в России, хотя в то время он считался там врагом народа номер один. Ведь Москва знала, что деньги он тратит на издание «Колокола». Можно ли вообразить что-либо подобное сегодня?⁷

В том же девятнадцатом номере «Арки» за 1987 год я дебютировал — под псевдонимом «Генрик Цивиньский» — весьма критической в плане концепции рецензией на книгу русского историка-эмигранта Михаила Геллера (1922–1997) *Машина и винтики. История формирования советского человека* (1985). Мой текст ждал публикации больше года, потому что в редакции его сначала попросту потеряли, а потом мои старшие коллеги очень бурно — о чем я узнал от Польковского — обсуждали, стоит ли вообще публиковать в «Арке» подобную рецензию. Дело в том, что еще в начале горбачевской «перестройки» я решительно оспорил тезис Геллера о том, что Советский Союз настолько преуспел в выковывании послушного *homo sovieticus*, что в наши 1980-е годы уже нет ни малейших шансов на какие-либо позитивные изменения в СССР:

⁷ *Rozmowa z Gustawem Herlingiem-Grudzińskim*, rozmawiał B. Sonik, „Arka” 1987, № 19, с. 4.

Не могу не отметить, что эта книга написана Геллером как бы от лица победителей — тоталитарная машина и винтики в описании автора взаимодействуют настолько идеально, что это могло присниться советским лидерам только в самых смелых снах...⁸.

Тогда я еще не знал, что столь же критически по отношению к этой максималистской книге Геллера был настроен его польский друг Густав Герлинг-Грудзиньский. Не соглашаясь с крайне пессимистическим видением человека и советского общества (*homo sovieticus*), которое можно было обнаружить еще в 1984 году в книге Геллера *Польша глазами Москвы*⁹, Герлинг увидел в этом конкретном друге-москале «живое доказательство» неэффективности и даже неустойчивости советской системы, заметив 23 сентября 1984 года в своем *Дневнике, писавшемся ночью*:

Геллер (в отличие от меня) серьезно относится к понятию *homo sovieticus* [...]. Однако из всех русских «оттуда», которых я знаю лично, он сам является, пожалуй, лучшим живым доказательством неустойчивости и поверхностности черт «советского человека». Кроме одной, которую система, похоже, сумела крепко-накрепко привить своим нынешним и бывшим подданным. Она выражается в непоколебимом убеждении, что никаких радикальных изменений «там» произойти не может. Это убеждение мне чуждо, я никогда не смогу с ним примириться¹⁰.

И уже 13 апреля 1985 года, вскоре после публикации французского перевода *Машины и винтиков (La Machine et les rouages)*, Герлинг продолжает:

Вне всяких сомнений, [книга Геллера. — Г.П.] очень хороша, она необычайно плотно, буквально на каждой странице, документирует попытки создания «советского человека», она в разы точнее «научных тезисов» Зиновьева. И все же Геллер в конце концов повторяет ошибку Зиновьева: принимает непрерывные попытки, неустанные усилия по созданию *homos'a* за доказательство того, что он практически создан (практически, ибо еще не в чистом виде)¹¹.

Уже в ноябре 1988 года я тайно переправлял через польско-чешскую и чешско-немецкую границы вышеупомянутый три-

⁸ H. Cywiński, *Historia pewnej utopii*, „Arka 1987, № 19, с. 144.

⁹ M. Heller, *Polska w oczach Moskwy*, Instytut Literacki, Paryż 1984.

¹⁰ G. Herling-Grudziński, *Dziennik pisany nocą*, vol. 2. 1982–1992, ред. W. Bolecki, [w:] G. Herling-Grudziński, *Dzieła zebrane*, ред. W. Boleckie, t. 8, Wydawnictwo Literackie, Kraków 2018, с. 149.

¹¹ Там же, с. 176.

надцатый номер «Арки» в Париж — Ирине Иловайской-Альберти. Я ехал на поезде из Катовице через Прагу в Мюнхен на годичную стипендию, и на чешско-немецкой границе таможенница — разумеется, чешская — вытащила из моего рюкзака журнал, быстро и цепко его просмотрела, а затем, о счастье, вернула владельцу... Сегодня-то мы знаем, что либо «агент Малешка» не считал нужным по горячим следам информировать КГБ о наших вдвойне секретных подвигах, либо КГБ не усмотрел в «русских» публикациях «Арки» какой-то особой угрозы для СССР... Добавим, не кто иной, как тот же Малешка — на сей раз под криптонимом «М.Г.» — вскоре опубликовал на страницах восемнадцатого номера «Арки» (1987) такую рецензию на тот тринадцатый номер:

В начале года появилось оригинальное, заслуживающее внимания издание — без указания даты и места публикации, без сведений об издателе, без цены, без названия. Разошлось оно успешно и скоро [...]. Тщательно сработанная обложка привлекает взгляд кириллическими буквами, образующими название русского эмигрантского журнала — «Русская мысль» — заголовки статей с первых страниц журнала за 1984 год: Сахаров, Польша, процесс над Комитетом защиты рабочих... Да-да! Буквы скорее привлекают, нежели отталкивают — особенно сегодня, когда с востока, похоже, опять доносится столько волнующих запахов (не обязательно радиоактивных), в которых некоторые надеются уловить пыль саморазрушающегося Кремля¹².

Добавим еще, что «отцом» этого русского, тринадцатого номера «Арки» на самом деле был парижанин Войцех Сикора, молодой краковский эмигрант (с 1981 года), в то время являвшийся зарубежным представителем журнала в столице Франции. Будучи человеком, чрезвычайно преданным делу и в своих действиях успешным, он добился укрепления связей между «Аркой» и эмигрантским еженедельником «Русская мысль», возглавляемым Иловайской-Альберти и неукротимой Наташей Горбаневской. В своей вступительной статье Иловайская написала:

Среди наших изгнанников и эмигрантов есть немало людей, которые, подобно сотрудникам нашей редакции и мне, кропотливо и терпеливо рабо-

¹² М.Г., „Rosyjska Myśl” i polski czytelnik, „Arka” 1987, № 18, с. 140. В подтексте этой легкой иронии — конечно, начатая Горбачевым в 1985 году «перестройка» и авария 26 апреля 1986 года на Чернобыльской АЭС. В целом же польский автор весьма доброжелательно относился к этому адресованному полякам голосу русской эмиграции.

тают над возрождением в русском обществе того духа, который царил в нем до того, как был убит ленинским переворотом. Мы боремся за уничтожение среди наших соотечественников всяческого расизма, фанатизма, нетерпимости, шовинизма, склонности к насилию, получивших широкое распространение за 70 трагических лет советской власти. Я считаю, что нет ничего более жестокого, чем отвергать и осуждать человека только на основании его национальности или вероисповедания. [...] Надеюсь, что все сказанное мною здесь не будет воспринято как попытка приуменьшить проблему исторической ответственности России перед Польшей. Уже говорила и повторю еще раз: я, родившаяся в эмиграции и всю жизнь прожившая на Западе, прихожанка Католической церкви, чувствую глубокую личную ответственность и всегда буду бороться против любого проявления гипотетического (бесперспективного) российского империализма. Пока я вижу только советский империализм — коммунистическую агрессию, направленную против человека. Я замечаю в советской власти стремление сформировать то, что принято называть, используя противоречивый термин, национал-большевизмом, но что на самом деле является попросту манипуляцией наиболее низменными человеческими инстинктами¹³.

Сегодня, конечно, уже очевиден догматизм столь добродушной аргументации, поскольку Иловайская-Альберти — кстати, немало послужившая русско-польскому делу — в смысле отношения большевизма к национальной традиции России, слишком явно пошла по стопам Александра Солженицына (1918–2008), который до конца жизни был убежден, что большевизм отрицал абсолютно всю прежнюю российскую историю и традицию. Что же касается обращения большевизма к низшим стадным инстинктам человека, то здесь Иловайская, вероятно, опиралась на идеи Александра Зиновьева (1922–2006), изложенные в его сборниках эссе *Гомо советикус* и *Мы и Запад* (оба опубликованы в 1982 году). Тем не менее, именно «Русская Мысль», возглавляемая Иловайской и Горбаневской, в восьмидесятые годы прошлого столетия много сделала для сближения русских и поляков, основанного — что крайне важно — на обоюдной правде.

Впрочем, М.Г. справедливо пишет, что парижский еженедельник «Русская мысль» ближе к национально-христианской версии и Солженицыну, нежели к Синявскому и «левачествующим критикам солженицынского идеала России» из «последней предгорбачевской группы эмигрантов...»¹⁴. Однако об Андрее

¹³ I. Iłowajska-Alberti, *Uwagi wstępne redaktora naczelnego „Russkiej Myśli”*, „Arka” 1986, № 13, с. 3. В издании, напоминаю, отсутствовали какие бы то ни было библиографические данные, формально это — публикация без даты и названия.

¹⁴ Там же, с. 141.

Синявском (1927–1997) здесь следует сказать подробнее. Знаменитый судебный процесс в середине февраля 1966 года в Москве, когда коммунистические власти судили его и Юлия Даниэля (1925–1988), вызвал огромный резонанс во всем цивилизованном мире. Обоих литераторов признали виновными и вынесли приговор за то, что они «нелегально», на Западе, издали под псевдонимами Абрам Терц и Николай Аржак свои гротескные и сатирические по отношению к СССР произведения. Кстати, впервые они были опубликованы Ежи Гедройцем (1906–2000) в парижском «Литературном институте»¹⁵.

Стойкие подсудимые во время этого беспрецедентного показательного процесса не проявили ни малейшего раскаяния и в отместку были приговорены к семи (Синявский-Терц) и пяти (Даниэль-Аржак) годам заключения в лагерях строгого режима. Вскоре другие диссиденты во главе с негибимым Александром Гинзбургом (1936–2002) составили *Белую книгу* процесса, объемом более четырехсот страниц, которая была распространена в СССР в «самиздате» и довольно быстро опубликована на Западе¹⁶. К со-

¹⁵ Даже и сегодня русские исследователи, говоря о процессе Синявского и Даниэля, очень редко, к сожалению, вспоминают, что книги Терца: *Суд идет* (1959), *Фантастические повести* (1961), *Любимов* (1963) и *Мысли врасплох* (1965), а также *Искушение* Аржака (1965) впервые увидели свет в Париже, на польском языке, именно у Ежи Гердройца. Поэтому здесь следует с благодарностью вспомнить Людмилу Сергееву, которая совсем недавно напомнила об этом факте. См.: Л. Сергеева, *Триумvirат // Л. Сергеева, Жизнь оказалась длинной*, АСТ, Москва 2019, с. 299. Однако она ошибается, утверждая, что публикация появилась в русскоязычном номере парижской «Культуры». Да, в 1960 году *Суд идет* Терца вышел у Гердройца по-русски — в первом из трех русскоязычных номеров «Культуры». Однако книжное издание этого произведения — подчеркнем еще раз, на польском языке — вышло годом ранее, это вообще была первая книга Терца на Западе. Густав Герлинг-Грудзинский, автор предисловия, написал: «Об авторе рассказа *Суд идет*, использовавшем псевдоним Абрам Терц, мы знаем только то, что это молодой советский писатель. Рукопись его новеллы (а точнее, небольшой повести) была передана в Париж и впервые печатается в польском переводе. В дальнейшем, очевидно, появятся переводы на другие языки». G. Herling-Grudziński, *Przedmowa // A. Terc, Sqd idzie*, Anonim, *Co to jest realizm socjalistyczny?*, пер. J. Łobodowski, Instytut Literacki, Paryż 1959, с. 7.

¹⁶ «Белая книга» о деле писателей Андрея Синявского и Юлия Даниэля, Possev-Verlav, Frankfurt/Main 1967. Следует добавить, что и Гердройц опубликовал — также до русскоязычной публикации — насчитывавшее сто с лишним страниц описание процесса с предисловием Элен Замойской, которая, кстати, мужественно переправила рукописи из СССР на Запад. Ср. *Sqd*

жалению, в то время у меня не было доступа к этому чрезвычайно важному изданию, но я узнал обо всем от безотказного Анджея Дравича, поскольку с октября 1981 года он уже преподавал в моем родном Институте русской филологии Ягеллонского университета. У своего научного руководителя, профессора Рышарда Лужного (1927–1998), я позаимствовал тогда очень полезную *Вольную русскую литературу* Юрия Мальцева (1931–1992), также, разумеется, изданную в эмиграции и содержащую краткий отчет о процессе над Синявским и Даниэлем¹⁷. Андрей Донатович Синявский, отбывавший свой суровый срок в Мордовской автономной области, был досрочно освобожден в 1971 году, а еще через два года получил долгожданное разрешение покинуть СССР. Вместе с женой Марией Васильевной Розановой он жил в Париже, а затем под Парижем, преподавал русскую литературу в Сорбонне. В 1975–1976 годах Синявский опубликовал знаменитые *Прогулки с Пушкиным* и *В тени Гоголя*, которые (особенно первую книгу) львиная доля русской эмиграции сочла «порочащими национальные святыни». В 1978 году, после известного конфликта с редактором «Континента» Владимиром Максимовым (1930–1995), Синявский с женой начали издавать журнал «Синтаксис», либеральный по своему мировоззрению и политическим оценкам и весьма задиристый по отношению к окостеневшей литературной традиции. В 1982 году Синявский опубликовал книгу о русском мыслителе и писателе Серебряного века Василии Розанове (1956–1918), а в 1984 году — увлекательные беллетризованные мемуары, озаглавленные *Спокойной ночи* — Дравич даже опубликовал на них в подпольной «Арке» короткую рецензию, где справедливо указывал на родство Синявского-Терца с Розановым:

Сознательно ли Синявский подверг переоценке свои возможности или поддался неумолимому наказу, во всяком случае, он пошел своим собственным путем синтеза: в направлении свободной эссеистики, автобиографической и критической рефлексии, промежуточным, межжанровым и богатым формам, которые кто-то назвал «литературой о литературе». Он следовал за великим русским мыслителем и писателем рубежа веков Василием Розановым, который как раз освободил русское письменное слово от бремени дидактики, обучая соотечественников свободе дискуссии¹⁸.

idzie! Stenogram z procesu A. Siniawskiego i J. Daniela (A. Terca i M. Arzaka). Moskwa, Luty 1966, Instytut Literacki, Paryż 1966.

¹⁷ Ю. Мальцев, *Вольная русская литература 1955–1975*, Possev-Verlav, Frankfurt/Main 1967, с. 55–56.

¹⁸ А. Drawicz, *Żyj długo i bądź sobą Abramie Tercu*, „Arka” 1987, № 17, с. 11.

Дравич лучше других ориентировался в идеологическом архипелаге и адресах русской эмиграции, а потому перед поездкой в Париж — в том январе-феврале 1987 года — снабдил меня всей необходимой информацией и соответствующими рекомендательными письмами к главным представителям двух вышеупомянутых течений русской эмиграции на берегах Сены. То есть, с одной стороны, направил меня к Иловайской, Горбаневской и через них к Максиму, а с другой — дал телефон и адрес Синявского и его энергичной супруги Розановой, которая вместе с мужем выпускала в Фонтене-о-Роз под Парижем «маленький, но дерзкий и заметный журнал “Синтаксис”»¹⁹. Журнал, добавим, оппозиционный по отношению не только к коммунизму, но и к любым национально-христианским идеологиям, даже, как в случае с «Русской мыслью» Иловайской и «Континентом» Максимова, откровенно экуменическими.

О своих парижских встречах с Иловайской и Горбаневской я потом писал несколько раз по разным поводам, а кроме того взял у них несколько интервью — и тогда, и во время своих последующих многочисленных посещений русского Парижа, а также в Кракове — для подпольной «Арки»²⁰, позже для газеты

¹⁹ Там же.

²⁰ *Rozmowa z Natalią Gorbaniewską*, „Arka” 1987, № 20, с. 46–55. *Rozmowa z Iriną Iłowajską-Alberti*, redaktorem naczelnym tygodnika „Russkaja Mysl”, „Arka” 1989, № 25, с. 64–71. Обе беседы подписаны уже упоминавшимся выше псевдонимом — Генрик Цивиньский. В первой Наташа Горбаневская высказала множество теплых слов о Викторе Ворошильском, прекрасном прозаике, поэте и переводчике, который узнал об этой подпольной публикации в марте 1988 года от своего товарища, также значимого поэта и переводчика Артура Мендзыжецкого (1922–1996). При этом произошло забавное — во всяком случае, с моей точки зрения — *qui pro quo*, о котором сегодня можно прочитать в третьем томе только что изданных «Дневников» Ворошильского: «Артур говорил об интервью Наташи в »Арке« с теплыми словами в мой адрес, кажется, его взял у нее Богдан Цивиньский, что меня удивило. Потом я нашел этот текст (“Арка” номер 20), Наташа разговаривала с каким-то молодым человеком, подписавшимся Генриком Цивиньским». W. Woroszyński, *Dzienniki*, t. 3: 1988–1986, Ośrodek Karta, Warszawa 2019, с. 31. Добавим на всякий случай, что настоящий Богдан Цивиньский (р. в 1939) — выдающийся польский историк и историк идей, подлинный знаток России, Украины, Белоруссии, Литвы. В 2015 он был отмечен жюри Премии им. Ежи Гедройца за книгу *Шанцы Культур. Эссе об истории народов Восточной Европы*. В настоящее время он ведет очень интересный блог *На взгляд старика*, <https://bohhdancywinski.pl/> (11.11.2021). В частности, отстаивает идею связи между империализмом царской России и СССР, предостерегая при этом поляков от формирования стереотипа «русского

«Час Краковски» и снова «Арки», уже вполне легальной²¹. А вот встреча с Синявским и Розановой в феврале 1987 года, к сожалению, не была осмыслена вплоть до конца 1993 года, не говоря уже о каких-либо публикациях. Поэтому я попытаюсь частично компенсировать это теперь, когда Андрея Синявского давно уже нет с нами, но, к счастью, еще живет в Фонтене-о-Роз под Парижем Мария Розанова.

Я приехал к Синявским 26 февраля 1987 года, в очень холодную, по французским меркам, зиму, которую даже называли тогда «зимой столетия». Выйдя из поезда RER (*Réseau Express Régional Île-de-France*), я с ужасом понял, что забыл адрес Синявского в парижской квартире на бульваре Шапель (метро «Сталинград»), где тогда жил. Поскольку 28 февраля я должен был окончательно вернуться в Польшу, другой возможности осуществить мечту о встрече с Синявским-Терцем в его доме под Парижем могло не представиться. Однако первая же проходившая мимо француженка средних лет при упоминании фамилии «Синявский» расплылась в улыбке: «*Rue Boris Vildé, huit*». И кивнула головой, указывая направление. Неделий ранее весьма сведущие польские и русские эмигранты предсказывали, что поговорить с самим Синявским у меня не получится, так как Мария Васильевна — «императрица», а потому немедленно возьмет бразды правления в свои руки и ни за что их не отдаст. Однако случилось так, что спутница жизни Андрея Донатовича спешно готовила новую публикацию и ушла на полтора часа, оставив меня один на один с писателем в его кабинете.

врага»: «Советский Союз продолжил четырехсотлетнюю традицию царской политики. Официально знание об этой преемственности исторического процесса было, разумеется, недоступно для моего поколения, однако кое-что подсказывала древняя польская литература, кое-что можно было понять, наблюдая за событиями дня сегодняшнего. Это не было насаждением мифической фигуры русского-врага, но лишь указанием на все еще присутствующую опасность, грозящую полякам со стороны русской империи, которая может называться по-разному, но реализует одну и ту же враждебную нам политику». <https://bohancywinski.pl/rosja-carow-rosja-bolszewikow/> (31.11.2021).

²¹ *Rosjanie i Polacy są podobni*, z Natalią Gorbaniewską rozmawiają G. Przebinda („Arka”) i A. Nowak („Czas”), opr. J. Korzeniewska, „Czas Krakowski” 1990, 6–7 X, № 234, с. 36 <http://www.portal.arcana.pl/Gorbaniewska-rosjanie-i-polacy-sa-podobni,4144.html> (31.07.2021). *Rosję uratują młodzi Rosjanie*, z Iriną Howajską-Alberti rozmawia G. Przebinda, „Arka” 1992, № 39–40, с. 122–127.

В то время я мало что знал о Синявском²², но все это были вещи по-своему примечательные, связанные с яркими биографическими и литературными ассоциациями. Поэтому я смог сообщить писателю, что его задержание и арест 8 сентября 1965 года произошли как раз в том месте — у Никитских ворот на Бульварном кольце в центре Москвы, — которое Булгаков в романе *Мастер и Маргарита* вписывает в путь Ивана Бездомного, гнавшегося за Воландом. Затем Иван Бездомный преследовал таинственного заграничного профессора по той же улице Герцена (сегодня Большая Никитская), по которой двое реальных, из плоти и крови, кагебешника вместе с задержанным Синявским молнией пронесли на машине по направлению к бессмертной Лубянке. Я мог бы еще добавить, что и по сей день ясно вижу сходство литературного Абрама Терца с булгаковскими Коровьевым и Азazelло одновременно. В конце концов, в упомянутой выше повести *Спокойной ночи* Терц изображен следующим образом:

Я его как сейчас вижу, налетчика, картежника, сукиного сына, руки в брюки, в усиках ниточкой, в приплюснутой, до бровей, кепке, проносающего легкой, немного виляющей походкой, с нежными междометиями непристойного свойства на пересохших устах, свое тощее, отточенное в многолетних полемиках и стилистических разноречиях тело. Подобранный, непререкаемый. Чуть что — зарежет. Украдет. Сдохнет, но не выдаст. Деловой человек...²³.

Однако мы тогда говорили не о Булгакове²⁴, потому что Хозяин, разумеется, предстал передо мной в обличье не Абрама Тер-

²² Я однако кое-что знал о поэтике ранних произведений Терца, благодаря скрупулезному исследованию Александра Вата, выразительно озаглавленному *Читая Терца* (1961), с которым познакомился, благодаря изданию текстов Вата *Эвокация* в подпольном краковском издательстве KOS (1981). Эта серьезная семидесятистраничная статья первоначально была напечатана в качестве предисловия к *Фантастическим повестям* Терца, изданным Гедройцем в 1961 году.

²³ А. Терц, *Спокойной ночи. Роман*, Синтаксис, Париж 1984, с. 17.

²⁴ Я тогда находился под впечатлением подготовки текста и комментариев к *Мастеру и Маргарите*, еще в переводе Ирены Левандовской и Витольда Домбровского, но уже с предисловием Анджея Дравича. Публикация после перипетий с цензурой — главным образом, как мне кажется, из-за имени автора предисловия — появилась только в 1990 г. во II серии Национальной библиотеки. Синявский, конечно, не мог проглядеть Булгакова, но то, что он о нем говорил и писал, до сих пор вызывает во мне глубокий протест. Хотя бы такое абсурдное соображение, брошенное мимоходом в Варшаве в декабре 1993 года, в разговоре с критиком Славомиром Мазуреком: «Культурные немцы — хотя русских считали варварами — начинают верить, что ради

ца, но ученого-филолога — профессора Андрея Синявского. Так случилось, что незадолго до отъезда в Париж я изучал в Кракове пространный труд Николая Федорова (1829–1903) *Философия общего дела*. Этот весьма загадочный русский мыслитель-практик провозглашал там ни больше ни меньше как долг «сыновей» немедленно воскресить всех умерших «отцов», полагая, что это может быть осуществлено путем сексуального аскетизма и молитвы, а также использования последних достижений техники. При этом «дочерям» Федоров настоятельно не рекомендовал рожать детей, дабы не способствовать непрерывному движению «круговорота смерти». В своем тексте я критически рассмотрел эту достаточно странную мысль в контексте близкой мне универсальной философии Владимира Соловьева (1853–1900)²⁵. Однако Синявский дал понять, что сам он однозначно на стороне «безумного проекта» Федорова. Позже он выразил это в беседе со Славомиром Мазуреком в Варшаве:

Особенностью русской философии является способность ставить проблемы чрезвычайно смело и парадоксально. Так, как до сих пор не осмеливался сделать никто в мире. Возьмем, к примеру, такого автора, как Николай Федоров. В своей книге под названием *Философия общего дела* он провозглашает, что все человечество должно выполнить универсальную задачу. Задача эта описана в Евангелии — речь идет о воскрешении мертвых. Это должен сделать сам человек. Звучит, как фантазия чистой воды. (И, возможно, на самом деле является фантазией, хотя писал это человек глубоко верующий). Но под пером Федорова эта чисто богословская проблема, проблема воскресения, приобретает какую-то невероятную актуальность. Как будто здесь и сейчас, не ожидая нового знамения от Бога — все знамения уже были явлены — следует взяться за дело и стремиться к воскресению мертвых. Что ж, мои дорогие — а мои дорогие — это все человечество — у вас есть конкретная задача — вы должны воскресить своих умерших предков... Сперва поколение отцов... Создать соответствующую технологию... Ни один автор science fiction до такого бы не додумался. Это невероятно освежает наше мышление. Русские философы по-новому решают универсальные проблемы²⁶.

спасения расы, а может, и всего мира, нужно уничтожить евреев. Они хотят спасти мир, убивая детей в газовых камерах. Чисто иррациональная, дьявольская идея. *Мастер и Маргарита* — дитя поистине двадцатого века — миром тут правит сатана». A. Siniawski, *Rosja: przekłete problemy*, rozmawiał C. Mazurek, „Ex Libris” (dodatek do „Życia Warszawy”) 1993, декабрь, № 41, с. 2. Я еще вернусь к этому сюжету при описании моей второй встречи с Синявским, в Женеве в июне 1988.

²⁵ Por. G. Przebinda, *Włodzimierz Sołowjow i Mikołaj Fiodorow. Dwie koncepcje eschatologiczne*, „Slavia Orientalis” 1989, № 3–4, с. 611–622.

²⁶ A. Siniawski, *Rosja: przekłete problemy...*, с. 2.

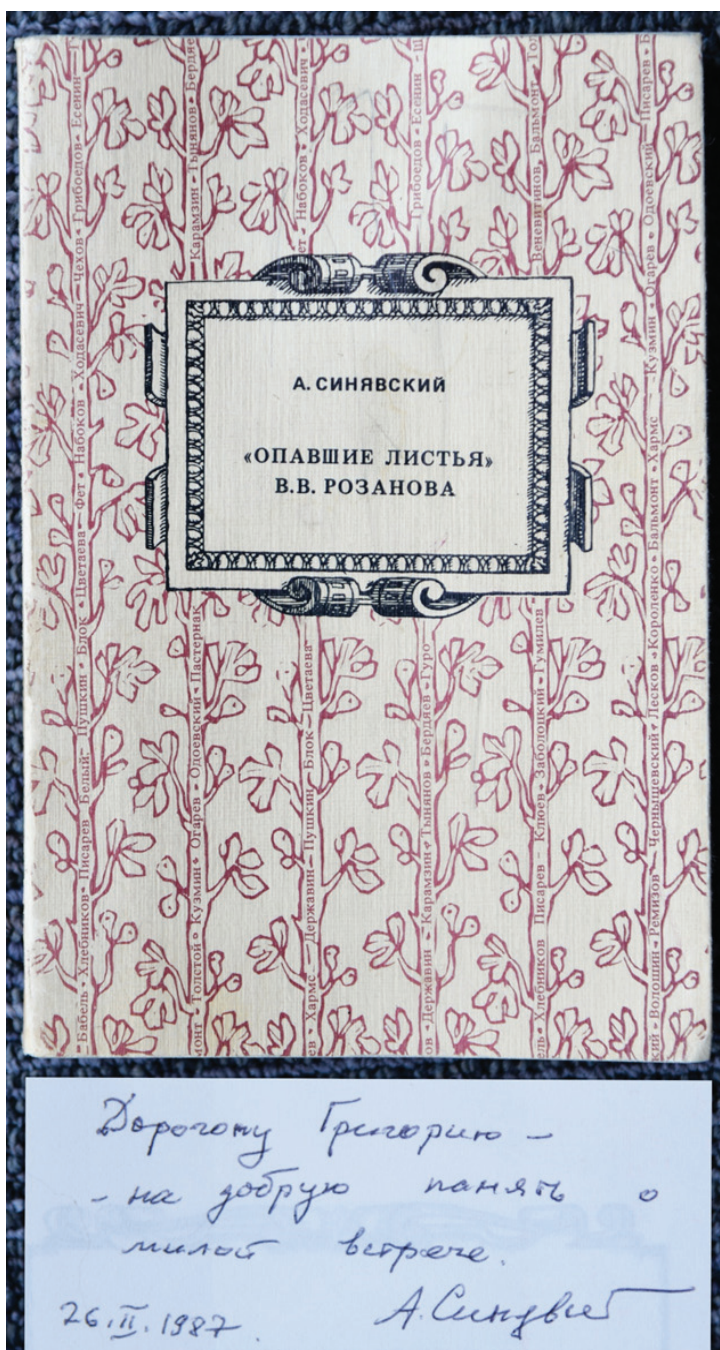
ТРИ ВСТРЕЧИ С АНДРЕЕМ СИНЯВСКИМ

Я же в феврале 1987 года в доме Синявского с большим вниманием слушал его литературоведческий монолог об Алексее Ремизове (1877–1957), великом эмигрантском писателе России первой половины XX века, и еще более обширную лекцию о Розанове, о котором Дравич уже говорил в связи с Синявским. В качестве прощального подарка я получил также упомянутые выше «Опавшие листья» Василия Розанова с теплой дарственной надписью на русском языке. К сожалению, мне не удалось сразу переправить эту ценную книгу из Парижа в Краков, так как во время обратного перелета, уже в варшавском аэропорту Окенче, меня подвергли тщательному досмотру и отобрали практически все русские и польские книги (всего более девяноста наименований). Увидев дарственную надпись Синявского таможенница иронически заметила: «Вот что у вас за знакомые...». Хотя мне трудно поверить, чтобы она что-либо знала о Синявском. Справедливости ради добавлю, что мне все же разрешили забрать не только польско-французский разговорник, но и — я вспоминаю об этом с большой благодарностью — *Собрание сочинений* Солженицына на русском языке. После долгих перипетий и выигранного мною в конце концов в середине 1988 года процесса в Высшем административном суде против Главного таможенного управления я вернул себе бесценные издания, все до единого²⁷. А уникальный подарок Синявского — книгу о Розанове — бережно храню в своей библиотеке по сей день.

* * *

Моя вторая встреча с Андреем Синявским произошла вскоре после этого, в июне 1988 года, в Женеве, на посвященном тысячелетию крещения Руси международном научном colloquium, который организовал Жорж Нива, крупный знаток России. Помимо Андрея Синявского и Марии Розановой, в числе докладчиков были писатель-«деревенщик» Валентин Распутин (1937–2015), историк Марк Раев (1923–2008), итальянский славист Чезаре Дж. Де Микелис, польский русист из Ягеллонского университета Луциан Суханек, русский литературовед-эмигрант

²⁷ Por. G. Przebinda, *Cło czy cenzura?*, „Tygodnik Powszechny”, 27 XI 1988, № 48 (2057), с. 3–4.



Дарственная надпись Андрея Синявского Гжегожу Пшебинде на заглавном листе книги *Опавшие листья* Василия Розанова

Ефим Эткинд (1918–1999), русский филолог и семиотик Борис Успенский, православный священнослужитель и историк Церкви из Мюнхена о. Кирилл Фотиев (1928–1990), бывший диссидент, а ныне историк из Израиля Мелик Агурский (1933–1991), бывший диссидент и узник ГУАГа, а ныне эмигрант из Лозанны Анатолий Левитин-Краснов (1915–1991), французская славистка Вероника Лосская (1931–2018) и ее муж, французский богослов Николай Лосский (1929–2017). Имея счастье быть приглашенным профессором Нива на этот пир духа, я обзавелся тогда очень полезными знакомствами со славистами мирового масштаба. О. Кирилл Фотиев еще раньше, в Мюнхене, помог мне перевести на русский язык мой женевский доклад о восприятии идей Соловьева в Германии, а Марк Раев, с которым я познакомился уже в самой Женеве, вскоре опубликовал в Нью-Йорке увлекательную полемическую переписку Соловьева с поляками²⁸. У меня также завязались теплые отношения с Анатолием Левитиным-Красновым, антикоммунистом и приверженцем «христианского социализма», но зарождавшуюся дружбу прервала его трагическая смерть в Люцерне в апреле 1991 года. Знаменательно, что 21 августа 1991 года, во время известного коммунистического путча, в Москве скоропостижно скончался энергичный Мелик Агурский, который на той женевской конференции справедливо призывал меня в будущем обратить внимание не только на экуменический универсализм Соловьева, но и на его панъевропейский империализм, к сожалению, отдававший расизмом, который выражался в антиазиатском «панмонголизме». Агурский имел в виду превнесение Соловьевым действий германского императора Вильгельма II по отношению к китайцам, восставшим против экспансии Европы и России. За месяц до смерти, узнав о карательной экспедиции, отправленной в Китай императором Вильгельмом — в отместку за убийство немецкого посла в Пекине, — русский мыслитель написал стихотворение *Дракон (Зигфриду)* с такой концовкой, явно в духе средневековых крестовых походов:

²⁸ Por. G. Przebinda, *L'accueil fait à la pensée de Vladimir Solov'ev en Allemagne et dans les pays germanophones 1889–1953*, „Cahiers du Monde russe et soviétique” (Le numéro spécial: *Le Christianisme Russe: entre millénarisme d'hier et soif spirituelle d'aujourd'hui*), Paris 1988, vol. XXIX (3–4), Juillet–Decembre, c. 427–445. Por. *Забывтое письмо Владимира Соловьева*, публикация Г. Пшебинды, перев. З. Юрьевой и А. Тюрина, „Новый журнал”, Нью-Йорк 1989, кн. 174, с. 217–223.

Наследник меченосной рати!
 Ты верен знамени креста,
 Христов огонь в твоём булате,
 И речь грозящая свята²⁹.

Сам Синявский в Женеве, в том же июне 1987 года, как истинный мудрец, совершал «апофеоз книги», отсылая, в основном, к позднесредневековой традиции Московской Руси. Он апологетически напоминал, и это уже тогда показалось мне очень близким, акт милосердия Ульянии Осоргиной-Лазаревской, богатой помещицы того времени (ум. 1604), которая открыла крестьянам амбары своего мужа. Случилось это в 1601–1602 годах — при Борисе Годунове — во время страшного голода, когда в Московской Руси питались человеческой плотью. Синявский сказал:

Замечательный образец такого рода свидетельской литературы представляет собой *Повесть об Ульянии Осоргиной XVII века*. Она была написана сыном Ульянии — Дружиной Осоргиным [...]. Это очень простой и безыскусный рассказ, наполненный множеством бытовых и психологических подробностей. В частности, Ульяния — святая в миру — превратила свою жизнь в подвиг любви и милосердия. Тогда русской земле пришлось испытать страшный голод. Живя в богатом доме своего мужа, Ульяния вдруг проявила необыкновенную прозорливость. Родители мужа удивлялись, говоря, что когда не было голода, Ульяния столько не ела, как она сейчас. А Ульяния ссылалась на то, что недавно родила, и потому все время хочет есть. На самом деле она тайно подкармливала голодающих. В подлинности всех этих, точно и тонко подмеченных деталей не приходится сомневаться³⁰.

Докладчик, как я вижу сегодня, несомненно, идет здесь по стопам как либерального историка Василия Ключевского (1841–1911), автора *Добрых людей Древней Руси*, так и эмигрантского историка идей Георгия Федотова (1866–1951), который также искал в Киевской и Московской Руси подобную святость, подтверждаемую гуманными и милосердными поступками. Таким образом, в случае Синявского апофеоз книги являлся фактически апофеозом активного милосердия. Жаль только, что столь проницательный филолог и историк культуры не заметил, что в *Мастере и Маргарите* аналогичный акт милосердия совершает заглавная героиня, освобождая от вечных мук трагическую

²⁹ Вл. Соловьев, *Стихотворения и поэмы*, Советский писатель, Москва 1974, с. 136.

³⁰ А. Синявский, *Апофеоз книги*, „Cahiers du Monde russe et soviétique”, с. 295, www.persee.fr/issue/cmr_0008-0160_1988_num_29_3 (11.11.2021).

детоубийцу Фриду. При, заметим, явном одобрении Воланда, которому Синявский в упомянутом разговоре с Мазуреком несправедливо приписал покровительство над газовыми камерами.

А теперь давайте отвлечемся: расскажу, при каких удивительных обстоятельствах продолжилась моя вторая встреча с Андреем Донатовичем в июне 1988 года. Профессор Жорж Нива, организовавший юбилейную конференцию на факультете славянской филологии, которым он руководил в Женевском университете, пригласил участников на ужин в свой дом, находившийся всего в тринадцати километрах от Женевы, но уже по другую сторону границы — во Франции. Напомню, что я приехал в Женеву из Мюнхена, и документы на швейцарскую визу подавал в Баварии. Подумать о французской визе мне и в голову не пришло, однако профессор Нива заверил, что спокойно провезет нас через границу на своей машине, так как ездит туда и обратно несколько раз в неделю, и никто особенно за этим не следит. Это был мой первый и последний опыт перехода через «зеленую границу»: за рулем сидел профессор Нива, а попутчиками были супруги Синявские (которым, вероятно, виза не требовалась) и Мелик Агурский, который в 1974 году участвовал вместе с Александром Солженицыным в знаменитой самиздатовской публикации *Изпод глыб*, а теперь беспокоился, не поймут ли его, гражданина Израиля, на франко-швейцарской границе без визы. Но все прошло именно так, как торжественно пообещал хозяин, и мой разговор в машине с Андреем Донатовичем завершился еще одним приглашением в Фонтене-о-Роз, на сей раз непременно — как он подчеркнул — с ночевкой. Тогда мы и предположить не могли, что наша третья и последняя встреча состоится в Кракове, через два года после распада СССР.

Однако прежде чем это произошло, в январе 1989 года, вскоре после смерти Юлия Даниэля 30 декабря 1988 года, Синявский впервые после отъезда из России посетил Москву. Правда, советские власти позаботились о том, чтобы он не попал на похороны друга. Будучи либералом и центристом, Синявский с самого начала был убежденным сторонником Михаила Горбачева, точно так же, впрочем, как его друг Анджей Дравич. Это вызывало яростный протест, прежде всего, Владимира Максимова, Натальи Горбаневской и Ирины Иловайской-Альберти³¹.

³¹ Por. A. Michnik, *Stalin wiecznie żywy*, „Gazeta Wyborcza” 2013, 25 V, <https://wyborcza.pl/magazyn/7,124059,13975516,stanin-zabija-noca.html> (19.08.2021).

Другой великий эмигрант, Александр Солженицын, постоянно обвинял Синявского в космополитизме, звучали подобные обвинения после распада СССР и на его родине. Московская газета «Литературная Россия» даже объявила Синявского главным русофобом, предложив «поступить с ним так, как мусульмане поступают с Салманом Рушди»³².

* * *

На первую встречу с Андреем Синявским я — как уже было сказано — отправился в Париж, во второй раз мы встретились, можно сказать, на полпути, — в Женеве, а в третий и последний раз мне посчастливилось увидеться с писателем в своем родном Кракове, после того как Синявский и его супруга были приглашены профессором Луцианом Суханеком в Институт русской филологии Ягеллонского университета. Выступали гости в нашем Малом Падеревианум на аллеях Мицкевича, 11, и это стало достойным продолжением «Русских встреч», которые Суханек организовал еще во времена «Солидарности», в том числе и во время военного положения. Упомяну здесь очень нашумевшую, справедливо антисоветскую конференцию «Лики России», состоявшуюся в Ягеллонском университете 11 и 12 декабря 1981 года, буквально накануне введения военного положения, а также литературную конференцию, посвященную Борису Пастернаку (1890–1960) поздней осенью 1983 года. На последней смог присутствовать уже хорошо нам знакомый Жорж Нива.

Из тех далеких времен я вспоминаю также приезды в Краков с лекциями великих варшавских русистов Виктора Ворошильского (1927–1996) и Северина Поллака (1907–1987), а также профессора Юзефа Жичиньского (1948–2011), тогда еще не епископа. На улице Июньского манифеста, 8, рядом с Collegium Novum Ягеллонского университета, он увлеченно рассказывал нам о французской книге Алена Безансона *Интеллектуальные истоки ленинизма* (1987). В ноябре 1993 года Синявский, следует подчеркнуть, большую часть времени провел в Варшаве, на конференции «Литература и демократия. Безопасные и опасные связи», организованной во дворце Сташица Институтом лите-

³² A. Siniawski, *Kłęska zwycięstwa*, „Gazeta Wyborcza” 1993, 9 X, № 237, с. 9.

ратурных исследований Польской академии наук и редакцией ежемесячника «Рес Публика Нова» (среди участников были Чеслав Милош и словенский писатель и эссеист Драго Янчар)³³.

Замечательное предложение взять в Кракове интервью у Синявских исходило от еженедельника «Тыгодник Повшехны», а точнее — от моего коллеги, также по Ягеллонскому университету, Анджея Романовского. Он был чрезвычайно взволнован — разумеется, приятно — этой уникальной возможностью. Романовский, вероятно, знал, что мы собираемся продолжать добрую традицию русиста и переводчика Збигнева Подгурца (1935–2020), который еще в первой половине 1970-х годов провел для «Тыгодника» длинную серию «русских» бесед — в том числе с Владимиром Максимовым, Андреем Тарковским (1932–1986), Эрнстом Неизвестным (1926–2016), Булатом Окуджавой (1924–1997), Владимиром Войновичем (1932–2018) и упоминавшимся в самом начале статьи Александром Зиновьевым.

В заключение я хотел бы высказать еще несколько соображений. Во-первых, Мария Розанова — как станет ясно из интервью, опубликованного ниже, — полностью компенсировала свое непреднамеренное отсутствие во время моей первой беседы с Андреем Синявским в их доме в Фонтене-о-Роз. Во-вторых, следует помнить, что часто встречающийся в разговоре термин «Белый дом» не имеет никакого отношения к Вашингтону: это тогдашнее здание Верховного Совета РФ, конфликтовавшего с Ельциным (сегодня — здание Правительства РФ). И в-третьих, буквально на следующий день после появления в «Тыгоднике Повшехном» этого интервью мне позвонила возмущенная, никогда не терявшая бдительности Наташа Горбаневская, кото-

³³ Синявский однако говорил там, главным образом, не о заглавной теме, а о метафизических, религиозных, даже эсхатологических задачах литературы, отсылая, в частности, к великому роману Бориса Пастернака: «Ползком, одолевая ужасы, скажу: а может быть, когда-нибудь так и будет, и об этом, единственно об этом, твердит вся мировая словесность. Искусство всегда надежда. Не об этом ли бормотал Пастернак в романе *Доктор Живаго*? «...Искусство всегда, не переставая, занято двумя вещами. Оно неотступно размышляет о смерти и неотступно творит этим жизнь. Большое, истинное искусство, то, которое называется Откровением Иоанна, и то, которое его дописывает». [...] Как рыба об лед, искусство бьется о неокончателность речи. Мы пишем лишь потому, что самое лучше, самое прекрасное еще не написано». А. Синявский, *Почему мы все-таки пишем? Искусство бьется о неокончателность речи // „Хранить вечно“* (Андрей Синявский), Специальное приложение к „НГ“ 1998, февраль, № 1, с. 2.

рая придерживалась диаметрально противоположного мнения относительно войны Ельцина с Верховным Советом. Поэтому вскоре я взял отдельное интервью у Наташи, человека мне очень близкого, хоть в этом вопросе мы и разошлись — оно было опубликовано в «Тыгоднике Повсехном» в январе 1994 года³⁴.

Перевод Ирины Адельгейм



Книги Синаевского-Терца и Даниэля-Аржака из домашней библиотеки Гжегожа Пшебинды

REFERENCES

ARKA. *Wybór 10–12*. Kraków — 1985. *Eseistyka. Krytyka. Literatura* M. Chudecki, and W. Sikora. Eds. Paryż: Aktis, 1986.

«*Belaya kniga*» o dele pisateley Andrey Sinyavskogo i Yuliya Danielya. Frankfurt/Main: Possev-Verlag 1987 [«*Белая книга*» о деле писателей Андрея Синаевского и Юлия Даниэля. Frankfurt/Main: Possev-Verlag, 1987].

³⁴ Jak Polak z Rosjaninem... Rozmowa z Natalią Gorbaniewską, rozmawiał G. Przebinda, „Tygodnik Powszechny” 1994, 30 I, № 5 (2315), s. 5–6.

TRZY WSTRECI S ANDRZEEM SINIAVSKIM

- Cywiński, Henryk [Przebinda, Grzegorz]. "Historia pewnej utopii." *Arka* 1987, no. 19.
- Drawicz, Andrzej. *Inna Rosja*. Warszawa: NOWA, 1980.
- Drawicz, Andrzej. "Żyć długo i bądź sobą Abramie Tercu." *Arka* 1987, no. 17.
- Gertsen, Aleksandr Ivanovich. *Sobraniye sochinyeniy v tridsati tomakh*. Moskwa: Izdatel'stvo Akademii Nauk SSSR. 1954–1960, Vol. 5 (1955): *Pis'ma iz Frantsii i Italii 1847–1852*. Vol. 10 (1956): *Byloye i dumy 1852–1868, chast' V* [Герцен Александр Иванович. *Собрание сочинений в тридцати томах*. Москва: Издательство Академии наук СССР 1954–1960. Vol. 5 (1955): *Письма из Франции и Италии 1847–1852* (1955). Vol. 10 (1956): *Былое и думы 1852–1868, часть V*].
- Herling-Grudziński, Gustaw. *Dziela zebrane*. Bolecki, Włodzimierz. Ed. Kraków: Wydawnictwo Literackie 2009–2021. Vol. 8 (2018): *Dziennik pisany nocą* (vol. 2: 1982–1992).
- Howajska-Alberti, Irina. "Uwagi wstępne redaktora naczelnego 'Russkoj Mysli'." *Arka*, 1986, no 13.
- "Jak Polak z Rosjaninem... Rozmowa z Natalią Gorbaniewską." Rozmawiał Grzegorz Przebinda. *Tygodnik Powszechny*, 1994, 30 I, no 5 (2315).
- Leszkowski, Marek [Maleszka, Lesław]. "Stan wojenny na łamach 'Russkoj Mysli'." *Arka* 1985, no 11.
- Literatura i demokracja: bezpieczne i niebezpieczne związki*. Warszawa: Instytut Badań Literackich PAN 1995.
- Mal'tsev, Yuriy. *Vol'naya russkaya literatura 1955–1975*. Frankfurt/Main: Possev-Verlag, 1967 [Мальцев Юрий. *Вольная русская литература 1955–1975*. Frankfurt/Main: Possev-Verlag 1967].
- M.G. [Maleszka, Lesław]. "Rosyjska Mysl' " i polski czytelnik." *Arka*, 1987, no 18.
- Przebinda, Grzegorz. "Apokalipsa Włodzimierza Sołowjowa." *Znak*, 1986, no. 11–12: 384–385.
- Przebinda, Grzegorz. "L'accueil fait à la pensée de Vladimir Solov'ev en Allemagne et dans les pays germanophones 1889–1953." *Cahiers du Monde russe et soviétique*, 1988. Vol. 29 (3–4).
- Przebinda, Grzegorz. "Clo czy cenzura?" *Tygodnik Powszechny* 1988, 27 XI, no. 48 (2057).
- Przebinda, Grzegorz. "Włodzimierz Sołowjow i Mikołaj Fiodorow. Dwie koncepcje eschatologiczne." *Slavia Orientalis*, 1989, no 3–4.
- "Rozmowa z G. Herlingiem-Grudzińskim." Rozmawiał Bogusław Sonik. *Arka*, 1987, no 19.
- Sergeyeva, Lyudmila. "Triumvirat." Sergeyeva Lyudmila. *Zhizn' okazalas' dlinnoy*. Moskwa: AST 2019 [Сергеева, Людмила. "Триумвират." Сергеева Людмила. *Жизнь оказалась длинной*. Москва: AST 2019].
- Sąd idzie! Stenogram z procesu A. Siniawskiego i J. Daniela (A. Terca i M. Arzaka)*. Moskwa, Luty 1966. Paryż: Instytut Literacki 1966.
- Siniawski, Andrzej. "Kłęska zwycięstwa." Trans. I. Lewandowska. *Gazeta Wyborcza* 1993, 9 X, no 237.
- Siniawski, Andrzej. "Dlaczego piszemy." Trans. J. Darczewska. *Literatura i demokracja. Bezpieczne i niebezpieczne związki*. Warszawa: Wydawnictwo Instytutu Badań Literackich PAN 1995.
- Sinyavskiy, Andrey. "Apofeozy knigi." *Cahiers du Monde russe et soviétique* 1988. Vol. 29 (3–4) [Синявский, Андрей. "Апофеоз книги." *Cahiers du Monde russe et soviétique* 1988. Vol. 29 (3–4)].

- Sinyavskiy, Andrey. "Pochemu my vse-taki pishem? Iskustvo b'yet-sya o neokonchatel'nost' rechi." *Khranit' vечно* (Andrey Sinyavskiy). Spetsial'noye prilozheniye v NG 1998, fevral' [Синявский, Андрей. "Почему мы все-таки пишем? Искусство бьется о неокончателность речи." *Хранить вечно* (Андрей Синявский), Специальное приложение в НГ 1998, февраль, no 1].
- Szczukin, Wasilij. "Trudne zmagania z przeszłością (Powieść Abrama Terca *Dobranoc*). *Emigracja i tamizdat. Szkice o współczesnej prozie rosyjskiej*. Suchanek, Lucjan. Ed. Kraków: Universitas, 1993.
- Terc, Abram. *Sąd idzie*. Anonim. *Co to jest realizm socjalistyczny?*. Trans. J. Łobodowski. Paryż: Instytut Literacki, 1959.
- Terlecki, Ryszard. "Moja historia 'Arki'," *Arka*, 1990, no 30.
- Terts, Abram. *Spokoynoy nochi. Roman*. Parizh: Sintaksis 1984 [Терц, Абрам. *Спокойной ночи. Роман*. Париж: Синтаксис, 1984].
- Wat, Aleksander. "Czytając Terca." Wat, Aleksander. *Ewokacja*. Kraków: Wydawnictwo KOS, 1981.
- Woźniak, Anna. *Ułuda i cud. W świecie sztuki Andrieja Siniawskiego — Abrama Terca*. Lublin: Wydawnictwo KUL, 2004



GRZEGORZ PRZEBINDA

 ORCID <http://orcid.org/0000-0001-9581-0214>

Uniwersytet Jagielloński

ANDRZEJ ROMANOWSKI

 ORCID <http://orcid.org/0000-0002-6832-5800>

Uniwersytet Jagielloński

ЕСЛИ ГОСУДАРСТВО ПРЕСТУПАЕТ ЗАКОН...**ИНТЕРВЬЮ С МАРИЕЙ РОЗАНОВОЙ И АНДРЕЕМ СИНЯВСКИМ (НОЯБРЬ 1993 Г.)¹**

Гжегож Пшебинда, Анджей Романовский: При всех различиях, какие существуют между поляками и русскими, есть у нас, пожалуй, одна общая черта. Мы придаем большое значение роли писателя в обществе — писателя, воспринимаемого как совесть нации, совесть эпохи. Сохранится ли эта его функция в России и в дальнейшем или же она ушла в прошлое вместе скоммунизмом?

Мария Розанова: Думаю, что идея писателя как совести эпохи умерла в России месяц назад. Она была расстреляна в Белом доме². Меня очень удивило и возмутило то, что никто из пи-

¹ Первопечатать (без комментариев): *Jeżeli państwo depce prawo... Rozmowa z Marią Rozanową i Andriejem Siniawskim* // «Tygodnik Powszechny» 1993, 21 X, № 47 (2315), с. 5. Польский вариант с комментариями: *Gdy państwo depce prawo... Z Marią Rozanową i Andriejem Siniawskim rozmawiają Grzegorz Przebinda i Andrzej Romanowski*, Kraków, listopad 1993 // «Studia Pigioniana» 2021, № 4. Кассета с записью интервью на русском не сохранилась, публикуется здесь в обратном переводе с польского, сделанным Ириной Адельгейм.

² Белый дом — неофициальное, но широко распространенное в России название здания на Москве-реке, где в 1991–1993 годах находился Верховный Совет Российской Федерации. Белый дом был обстрелян 4 октября 1993 года войсками президента Бориса Ельцина, которые вступили в вооруженное столкновение с крайне агрессивной и также использовавшей военные методы оппозицией во главе с вице-президентом Александром Руцким и председателем Верховного Совета Русланом Хасбулатовым. Начиная с 1994

сателей тогда не сказал: НЕТ! Это сделали лишь Виктор Розов и Юрий Поляков³.

И Андрей Синявский ...

М.Р.: — Да. И еще его «заклятый друг» — Владимир Максимов. Они с Синявским написали тогда, что за содеянное Ельцин должен каяться в монастыре⁴. В то время, как другие, лучшие люди Отечества, восприняли случившееся покорно. И даже заявили, что этого еще мало!⁵

А может, такая позиция объясняется переходным периодом — Россия отходит от коммунизма? То же самое происходило в начале советской власти, но со временем все изменилось.

М.Р.: — Но ведь ничего подобного событиям октября этого года в Москве никогда раньше не случалось. Да, в сталинские времена разыгрывались спектакли «судебных процессов», были «тройки»⁶, лагеря, беззаконие. Но меня всегда интересовало, за-

года Белый дом в Москве является резиденцией правительства Российской Федерации.

³ Известный драматург Виктор Розов и прозаик Юрий Поляков весьма критически оценили тогдашнюю позицию Ельцина. Об этом Поляков заявил в статье «Оппозиция умерла. Да здравствует оппозиция!», опубликованной 7 октября 1991 года в газете «Комсомольская правда». В среде эмиграции действия Ельцина осудил Юрий Литвинов, участник знаменитой демонстрации 25 августа 1968 года на Красной площади против введения в Чехословакию войск стран Варшавского договора.

⁴ Все, кто тогда интересовался Россией, знали, что представители парижской эмиграции — главный редактор «Континента» Владимир Максимов и редактор «Синтаксиса» Андрей Синявский — находятся в состоянии острого конфликта. Однако во время столкновения Ельцина с Верховным Советом произошло их примирение. Совместное письмо Максимова, Синявского и Петра Егидеса было опубликовано 16 октября 1993 года в «Независимой газете». В этом тексте авторы призывали Ельцина подать в отставку и отправиться в монастырь на покаяние.

⁵ Осенью 1993 года подавляющее большинство российской демократической интеллигенции поддерживало Ельцина во всех его начинаниях, включая военные действия против Верховного Совета во главе с Хасбулатовым и Руцким, обвинявшимися в попытке реставрации коммунизма. Широкую огласку получило тогда опубликованное 5 октября в газете «Известия» «Письмо сорока двух», декларировавшее безоговорочную поддержку ельцинской кампании и подписанное, в частности, Булатом Окуджавой, Беллой Ахмадулиной, Виктором Астафьевым, Алесем Адамовичем, Даниилом Граниним, Дмитрием Лихачевым, Юрием Нагибиным, Юрием Давыдовым, Александром Кушнером, Робертом Рождественским, Василием Быковым, Григорием Баклановым, Анатолием Приставкиным, Борисом Васильевым.

⁶ «Тройка» — печально известные псевдо-суды НКВД, состоявшие из трех человек — в 1930-е годы по приказу Сталина они массово выносили суровые приговоры тем, кто обвинялся в действиях во вред государству.

чем в 1930-е годы из людей под пытками выбивали признание вины. И я пришла к выводу, что палачи стремились убедить самих себя, что наказывают действительно виновных. А Ельцин даже об этом не позаботился. Беззаконие произошло среди бела дня, в центре города. В Белом доме погибло около 1500 человек.

В Польше об этом молчат... Разоблачительную статью из «Независимой газеты» о количестве жертв перепечатадо только «Жиче Варшавы»⁷.

М.Р.: — Давным-давно в брошюре *Что такое социалистический реализм* Синявский заметил, что с советской властью у него расхождения чисто стилистические⁸. Вот и теперь у нас происходит нечто в этом роде. Говорить при помощи танков — вопрос стиля! И такой стиль неприемлем.

Вы опасаетесь психологических последствий для россиян ...

М.Р.: — Я опасуюсь вопиющего нарушения заповеди «Не убий». И опасуюсь нарушений закона. Ведь, насколько мне известно, глава государства более всех прочих обязан стоять на страже закона! *Пионер всем детям пример!* А сегодня ловкие юристы говорят: да что конституция, она же брежневская!.. Но простите: закон о президенте был принят на основе именно этой конституции. Ельцин его принял, и Верховный Совет тоже. И на основании именно этого закона Ельцин стал президентом⁹.

Почему новый режим пользуется такой большой поддержкой со стороны россиян? Поляки объясняют это очень просто: русским, мол, нравится сильная рука...

М.Р.: — Думаю, поляки правы. Хотя не следует забывать, что говоря так, я, возможно, не являюсь типичной русской. Моя

⁷ Н.Никитенко, *Сколько трупов было в Белом Доме?* „Независимая газета”, 30 X 1993. По сей день точно неизвестно, сколько человек погибло во время боев за Белый дом — предположительно от 150 до 400.

⁸ Статью «Что такое социалистический реализм?» Синявский написал в Москве в 1957 году. Впервые текст был опубликован анонимно на французском языке в 1959 году, в февральском номере парижского издания «Esprit» (*Qu'est-ce que le Réalisme Socialiste?*). В 1959 году он увидел свет также на польском языке — у Ежи Гедройца в Париже. — А. Terc, *Sqd idzie — Аноним, Co to jest realizm socjalistyczny?*, Instytut Literacki, Paryż 1959, пер. J. Łobodowski. Przedmowa: G. Herling-Grudziński. «Абрам Терц» — литературный псевдоним Андрея Синявского.

⁹ Вышеупомянутая конституция была принята Верховным Советом СССР 7 октября 1977 года. Конституция, действующая в России в настоящее время, была принята в результате всенародного референдума 12 декабря 1994 г., то есть вскоре после уже упоминавшихся кровавых событий в Москве.

прабабушка была полькой, она вышла замуж в России за выходца из семьи польских ссыльных после восстания 1863 года. Прабабушка никогда не говорила по-русски, она ненавидела этот язык. И даже моей бабушке приходилось разговаривать со своей матерью — этой самой прабабушкой — исключительно на польском, немецком и французском языках. Так что, возможно, во мне осталось что-то от их неприятия...

Андрей Донатович, как вы относились к Ельцину до 3 октября?

Андрей Синявский: — Я никогда не был от него в восторге. Конкретных достижений я не увидел — разве что отпущенные цены.

А в чем, на ваш взгляд, заключался конфликт между президентом и парламентом? Что отстаивали стороны этого конфликта?

А.С.: — На мой субъективный взгляд — это борьба за власть, не более того. Борьба коммунистов против коммунистов. То, что Ельцин сдал партбилет раньше других, что он вместе с Руцким¹⁰ пришел со свечкой в церковь, не означает, что они оба сделались христианами...

М.Р.: — Что касается нашего прежнего отношения к Ельцину... В августе 1991 года в Париже я начала готовить 31-й номер «Синтаксиса»¹¹. И вдруг путч¹² — напряженные дни, бесконечные

¹⁰ Александр Руцкой (р. 1947) — с июня 1991 г. вице-президент Российской Советской Федеративной Социалистической Республики при президенте Борисе Ельцине. Он активно поддерживал Ельцина во время августовского путча 1991 года, о котором речь пойдет ниже. Однако позже между Руцким и Ельциным возник очень острый конфликт, прежде всего связанный с экономической политикой премьер-министра Егора Гайдара. Осенью 1993 года Руцкой активно сражался против Ельцина вместе с враждебно настроенным Верховным советом во главе с Хасбулатовым. После поражения 4 октября 1991 года арестован и снят с должности вице-президента, которая, впрочем, вскоре была упразднена. 26 февраля 1994 года Руцкой в рамках амнистии вышел на свободу. В 2006–2010 годах избирался губернатором Курской области. С 2013 года является членом пропутинского «Комитета поддержки реформ президента России».

¹¹ Эмигрантский, выходящий нерегулярно ежеквартальник «Синтаксис», издаваемый Марией Розановой и Андреем Синявским в Фонтене-о-Роз под Парижем с середины 1978 по 2001 год (с 1983 по 2001 год главным редактором осталась Розанова). В общей сложности вышло тридцать семь номеров журнала, который своим названием отсылает к самиздатскому поэтическому альманаху «Синтаксис», который Александр Гинзбург выпускал с декабря 1959 года по апрель 1960 года в Москве (всего три номера).

¹² Речь идет о крайне неудачном «московском путче», осуществленном 19–21 августа 1991 года группой рьяных коммунистов во главе с вице-президентом

телефонные звонки из Москвы, наконец, известие о том, что демократия побеждает, что «вчера мы стали европейцами»¹³. Н На первой странице мы поместили лозунг: «Наше дело — левое, победа будет за нами». Это, разумеется, аллюзия на лозунг военных лет...

Да, знаменитый лозунг Молотова: «Наше дело правое, враг будет разбит, победа будет за нами»¹⁴.

М.Р.: — Совершенно верно. Так вот, подготовку номера я закончила в октябре. Но на последней странице внесла необходимую коррективу: «Наше дело (конечно), левое, (но, к сожалению) победа будет за НИМИ»...¹⁵ Три дня эйфории, переживаний и — как видите — не замедлившие последовать сомнения. Потому что эти победившие демократы оказались не демократами! Это было очевидно, по крайней мере, для меня, еще в октябре 1991 года. Стиль определяет всё. Чем свердловская мафия лучше мафии днепропетровской?¹⁶

Вы говорите, как Владимир Максимов. Недавно, будучи в Польше, он заметил, что в России не обойдется без кровопролития, что Ельцин и его команда — обыч-

СССР Геннадием Янаевым (1937–2010). Своей целью участники путча ставили свержение президента СССР Михаила Горбачева и восстановление мощи Советского Союза. Однако благодаря решительной позиции Бориса Ельцина (1931–2007), тогдашнего президента Российской Федерации, еще входившей в состав СССР, оказались молниеносно разгромлены. Эти события привели к распаду СССР в декабре 1991 года.

¹³ «Сегодня 22 августа. Утро началось с радостного голоса [по телефону] из Москвы: „Поздравляем, мы стали европейцами! Победа!“». — М. Розанова, *Переворот*, «Sintaxis» 1991, № 31, с. 189.

¹⁴ Это заключительная фраза обращения Вячеслава Молотова, тогдашнего заместителя председателя Совнаркома и министра иностранных дел СССР, к советскому народу по радио в воскресенье 22 июня 1941 года, в 12 часов дня, то есть через несколько часов после нападения Германии на СССР.

¹⁵ Мария Розанова написала: «На первой странице этого номера, в лозунге, в результате эйфории, в которой пребывала редакция после путча, допущена серьезная ошибка. Текст следует читать [теперь]: «Наше дело (конечно) левое, (но, к сожалению) победа будет за НИМИ». — «Sintaxis» 1991, №31, с. 188.

¹⁶ Под «свердловской мафией» Розанова подразумевает команду Ельцина, политически связанную с Свердловском (ныне Екатеринбург). Днепропетровской мафией на закате СССР называли людей генерального секретаря Леонида Брежнева, который родился под Екатеринославлем (ныне Украина), которых он переводил в Москву и назначал на очень высокие посты.

ная мафия¹⁷. И он, и вы, кажется, забываете, что Ельцин выбрал народ, что он не самозванец.

М.Р.: — Какой народ — такой и президент. И какой народ — такой парламент... Что же касается Максимова: если в «Истории КПСС» написано, что трава от дождя становится мокрой, я не буду возражать только потому, что это написано в «Истории КПСС»... Точно так же я не буду оспаривать ту или иную идею потому лишь, что ее высказывает Максимов. В этом случае он как раз прав.

А каково ваше отношение к Горбачеву? Разве — особенно в период между августом и декабрем 1991 года, между путчем Янаева и образованием Содружества Независимых Государств — Горбачев не заслуживал большей поддержки?

А.С.: — Я однажды назвал Горбачева «первым диссидентом Политбюро». На слово «Политбюро» «Русская мысль»¹⁸ не обратила внимания и отреагировала очень резко: «Как это он — диссидент? Диссиденты — мы!»

М.Р.: — Горбачев был интересной фигурой. А интересное мне, как правило, нравится. Худшая смерть — смерть от скуки. Горбачев, несмотря ни на что, сделал для страны очень много. Вот мы сейчас сидим с вами здесь, в Кракове, — спасибо дяде Горбачеву... В Польше больше нет российских войск — спасибо Горбачеву: он начал этот процесс, Ельцин только завершил его¹⁹. Советская армия покинула Афганистан — спасибо Горбачеву²⁰.

¹⁷ Владимир Максимов (1930–1995) посетил Польшу дважды, в 1993 году он был в Варшаве и в Кракове, а в 1994 году — только в Варшаве. О «ельцинской мафии» Максимов сказал в Ягеллонском университете, на отделении восточнославянской филологии, в середине 1993 года.

¹⁸ Расцвет эмигрантского еженедельника «Русская мысль», издающегося в Париже с 1947 года, относится к 1978–2000 годам и связан с именем Ирины Иловойской-Альберги. Во времена «Солидарности» газета — прежде всего, посредством текстов Натальи Горбаневской — поддерживала демократическое движение в Польше. В период же перестройки она была очень критически настроена по отношению к Михаилу Горбачеву, подобно почти всем советским эмигрантам (за исключением Синявских). Во время кровавого конфликта Ельцина с Верховным Советом в 1993 году «Русская мысль» заняла радикальную проельцинскую позицию.

¹⁹ Советские войска окончательно покинули Польшу 17 сентября 1993 года, в тот период, когда президентом Польши был Лех Валенса, а президентом России — Борис Ельцин.

²⁰ Советские войска вошли в Афганистан при Брежневe 25 декабря 1979 года и были выведены при Горбачеве 15 февраля 1989 года.

А Литва?²¹

М.Р.: — Да, этот аргумент часто приводят: а Литва? А Тбилиси?²² Знал об этом Горбачев или нет? Тут все не так просто. Литовцы всегда были готовы пойти на огромные жертвы, лишь бы вырваться из состава СССР. После войны в Литве существовали «лесные братья»²³, бывшие литовские подпольщики. Литовцы воевали, умирали... И что? При Горбачеве за свободу заплатили своей кровью тринадцать человек. Мне этих людей очень жаль — но давайте сравним события января 1991 года с потоками крови, пролитыми в Литве после 1940 года...

Дело не в количестве жертв. Горбачев запустил процесс распада империи, но не осознавал этого. Он хотел сохранить империю, поэтому на нем кровь тринадцати литовцев. Хотя сам он сегодня утверждает, что ничего не знал.

М.Р.: — Возможно, Горбачев не всегда понимал, что он на самом деле начал. Когда Синявский пишет роман — знает ли он, чем все закончится? Нет.

Но ответственность писателя отличается от ответственности политика.

М.Р.: — Так что я допускаю две возможности: или это происходило без ведома Горбачева, или Горбачев лжет. Но, пожалуйста, поймите. Горбачев работал в очень тяжелых условиях. Он боролся с собственным Политбюро, постепенно, шаг за шагом устраняя противников. Ведь он далеко не всё мог себе позволить! Всё мог себе у нас позволить только товарищ Сталин. Глядя на Горбачева сейчас, я думаю, что он просто действовал слишком быстро.

Мы в Польше полагали, что слишком медленно...

²¹ Задавая этот вопрос, мы имели в виду кровавые события в советской еще Литве: 13 января 1991 года четырнадцать литовцев были убиты и несколько сотен ранены во время штурма Вильнюсской телебашни военнослужащими ВС СССР.

²² Кровавые события в столице советской Грузии произошли в ночь с 8 на 9 апреля 1989 года. Войска Советского Закавказского военного округа подавили мирную демонстрацию грузин на проспекте Руставели в Тбилиси. В ходе кровопролитной операции шестнадцать грузинских демонстрантов были убиты военными, трое позже скончались в больнице и несколько сотен оказались ранены.

²³ «Лесными братьями» называли литовских, латышских и эстонских партизан, сражавшихся против Советской армии после присоединения прибалтийских государств к СССР в 1944 году. Наиболее тяжелые бои велись до 1947 года.

М.Р.: — Вот именно! В то время я и сама так думала...

Ваша критика Ельцина касается также Егора Гайдара. А, собственно, почему? Вы обвиняете его в том, что он не претворил в жизнь свои экономические планы или что он не захотел это сделать?²⁴

М.Р.: — Я обвиняю его в том, что он сделал. А именно — ограбил народ. И предоставил десяти процентам общества возможность жить за счет оставшихся девяноста.

В Польше тоже иногда так говорят: мол, Бальцеро-вич ограбил народ...

М.Р.: — Но прежде, чем утверждать подобное, нужно оценить падение уровня жизни в Польше. Потому что в России он всего за полтора с лишним года упал минимум в десять раз. Мы находимся на уровне конца войны. А ведь тогда был голод...

А.С.: — Гайдар не имел права проводить реформы таким образом. Он лишь отпустил цены, но не создал никаких возможностей для частного сектора.

М.Р.: — Понимаете, мы в России очень хорошо знаем, что в Польше жить сложно. Но для нас, россиян, Польша — экономический рай.

Но можно ли винить Гайдара в том, что он за полтора года не все успел сделать?

М.Р.: — Он мог бы перестроить налоговую систему так, чтобы было выгодно развивать частные предприятия. К сожалению, Гайдар считал, что государству нужны деньги, и боялся мафии директоров госпредприятий. В результате производство остановилось, и теперь вся псевдорыночная экономика терпит фиаско... Борьба за власть у нас — это борьба за деньги. И происходит это ценой катастрофического расслоения общества.

Андрей Донатович, вы писатель, глубоко укорененный в русской культуре, автор книг о древней Руси, о России рубежа XIX — XX веков. Помогают ли эти знания в области русской истории, культуры и литературы понять, что происходит сегодня?

А.С.: — Жить — безусловно помогает. Однако не способствует футурологическим размышлениям. Другое дело история. Бу-

²⁴ Егор Гайдар (1956–2009), в 1991–1992 годы исполнявший обязанности премьер-министра, вице-премьер, министр финансов и экономики, пытался осуществить в России либеральную реформу экономики, потерпел поражение вследствие противодействия консервативных сил и полной неготовности общества к переменам и жертвам.

дучи писателем, я воспринимаю настоящее скорее интуитивно. Кровавый обстрел Белого дома мне не нравится вовсе не потому, что я арифметически вычислил какие-то закономерности развития России...

Но месяц назад в своем интервью для «Газеты Выборчей» вы ведь сравнили состояние русской интеллигенции в 1993 и в 1917 году. Вы сказали: «Я слишком хорошо помню, как Россия утрачивала свободы, провозглашенные в 1917 году, как закрывались неудобные газеты, вводилась цензура, запрещалась деятельность оппозиционных партий, и как интеллигенция, моя любимая интеллигенция, все это оправдывала»²⁵...

А.С.: — Да, я поддерживаю эту аналогию. Скрепя сердце.

В последней книге — «Иван-дурак»²⁶ — отчетливо прослеживаются ваши религиозные корни. Между тем вы избегаете называть себя религиозным писателем. Почему?

А.С.: — Потому что я боюсь людей, которые говорят о себе: «Я религиозный писатель». Для меня писатель — частное лицо, выражающее свои личные взгляды. Он может быть верующим, может быть атеистом. Однако определяя себя как религиозного писателя, он — особенно в России — претендует на роль папы римского. Словно хочет диктовать, как людям следует жить. И тогда возникают соблазны, на которых не один писатель свернул себе шею. Приведу конкретный пример. Мой любимый Гоголь в конце жизни опубликовал *Выбранные места из переписки с друзьями*. Где начал вдруг вещать — каким должен быть барин, каким — царь²⁷. А ведь писателю никто не дает такого права. Или второй пример, из сегодняшнего дня: Солженицын. Он считает себя пророком. Совершенно серьезно утверждает,

²⁵ А. Siniawski, *Kłęska zwycięstwa*, „Gazeta Wyborcza”, 9 X 1993, s. 9, пер. I. Lewandowska.

²⁶ А. Синяевский, *Иван-дурак. Очерк русской народной веры*, Sintaxis, Париж 1991. Книга подготовлена на основе лекций, которые Синяевский читал студентам Сорбонны с конца 1970-х годов.

²⁷ Николай Гоголь (1809–1852) — как автор *Ревизора* (1836), *Мертвых душ*, *Шинели* (1842) — был любимым писателем оппозиционной интеллигенции в России. Однако после публикации в 1847 г. прогосударственных *Выбранных мест из переписки с друзьями* он подвергся суровой критике со стороны недавних апологетов.

что им руководит сам Господь!²⁸ Поэтому я думаю, что если кто-то считает себя религиозным писателем — это слишком большая ответственность. Либо человек действительно пророк, и тогда об этом нужно молчать, либо он лжепророк. Если писатель сам себя полагает «совестью эпохи», остается лишь вздохнуть. Но в России такое явление, к сожалению, — обычное дело. Среди великих русских писателей подобными амбициями не грешил, собственно, только Пушкин. Он говорил о себе: я всего лишь поэт.

Солженицын в недавнем интервью еженедельнику «Русская Мысль» сказал: «После событий 3–4 октября можно только выразить неуверенную надежду, что хоть теперь, может быть, народ станет жить хоть немного лучше»²⁹. Так что он говорит совсем не так, как вы...

А.С.: — Газету «Русская мысль» я сейчас читать не в состоянии. Хотя это единственная эмигрантская газета в Европе, ее демонстративно проельцинская позиция вызывает во мне решительный протест.

Что будет дальше — с нами и с Россией? Мы помним солидарность поляков с русскими во время августовского путча 1991 года. Сейчас всё иначе, поляки — и даже другие народы Центральной Европы, не обремененные, в отличие от нас, антироссийскими настроениями, — снова боятся России. Оправдан ли этот страх?

М.Р.: — Более чем. Если правительство не действует по закону, то никогда нельзя быть уверенным, что оно совершит через неделю. Если государство само преступает закон, то нет никаких гарантий, можно ожидать чего угодно.

«Мне сегодня тяжело: почти вся русская интеллигенция за Ельцина, а я — против. А это трудно и стыдно — быть одному». Это из вашего недавнего выступления³⁰. А вы тоже чувствуете себя одинокой?

М.Р.: — Мне никогда не нравилось большинство. Я хорошо помню, как большинство боготворило Сталина. Я уважаю

²⁸ Солженицын, разумеется, никогда этого не говорил. Синявский находился в остром конфликте с автором «Архипелага ГУЛАГа» и порой приписывал ему фантастические вещи. — Ср., например: А. Синявский, *Солженицын как устроитель нового единомыслия*, «Sintaxis» №14, 1985, с. 16–32.

²⁹ Александр Солженицын о России накануне выборов, «Русская мысль» 1993, 28 X — 3 XI, № 4002, с. 1.

³⁰ A. Siniawski, *Kłęska zwycięstwa*, «Gazeta Wyborcza», 9 X 1993, s. 9.

и ценю людей, имеющих смелость противостоять большинству.

А.С.: — Скажу иначе. Журналисты иногда спрашивают меня: может, Россия не создана для демократии? И я допускаю такую возможность. Но даже в этом случае интеллигенция — меньшинство — должна быть за демократию. Может, политика Ельцина и окажется эффективной, ну так что ж? Моя совесть в любом случае будет против его методов.

М.Р.: — За последние два года в России произошли две ужасные вещи. Во-первых, коммунисты, которые в августе 1991 года были просто смешны, сейчас окрепли. Во-вторых, люди начали ненавидеть само слово «демократия».

А.С.: — Ведь получается, что демократия может стрелять по парламенту. И интеллигенция это поддерживает. Это действительно компрометирует демократию и интеллигенцию.

Так куда же идет Россия?

А.С.: — Я не делаю прогнозов. В свое время я не сумел предсказать появление Горбачева. Я не думал, что Россия так легко отдаст Польшу и другие страны, что это произойдет бескровно. Я был тогда невероятно счастлив, потому что сбылось то, о чем я мечтал — еще в лагере...

М.Р.: — Надо помнить, что в лагере Синявский сидел с литовцами, латышами и эстонцами. У них были очень хорошие отношения, хотя он представлял, так сказать, нацию оккупантов. Но он всем им давал свободу. С легкостью отпускал из Союза...³¹

Через месяц в России состоятся парламентские выборы. Что-нибудь изменится?

А.С.: — Мне кажется, народ находится в состоянии апатии, политика ему опротивела. Ельцин только что расстрелял один парламент — и хочет избрать новый... Демократы, не демократы... Наплевать!³²

³¹ Андрей Синявский сидел в мордовских лагерях в 1966–1971 годах. Власти досрочно освободили его 8 июня 1971 года. С 1973 года он находился в эмиграции в Париже, где в 1974 году опубликовал мемуары под названием *Голос из хора*, основанные, главным образом, на письмах из лагеря к Марии Розановой.

³² Выборы в Первую Государственную Думу Российской Федерации состоялись 12 декабря 1993 года. Проельцинский «Выбор России» получил 64 места, столько же — профашистская Либерально-демократическая партия России Владимира Жириновского, третье место заняли коммунисты — 42 места. Однако формально на выборах победили люди Жириновского, набравшие

Значит, вы не видите надежды?

М.Р.: — Надежда существовала при Горбачеве. Мы шаг за шагом двигались вперед. В эпоху Ельцина надежда оказалась утрачена. Причем очень быстро.

Солженицын вернется в Россию в мае следующего года³³. А вы?

Что мне там делать? Да, я получил такое предложение, мне даже пообещали выборы в депутаты Верховного Совета. Но я только рассмеялся: зачем мне это? Я не думаю, что писатель должен быть депутатом. Возвращаться в Россию, прилагать усилия к тому, чтобы меня избрали? А потом, возможно, оказаться в таком же расстрелянном парламенте? Нет-нет. Во Франции я могу, по крайней мере, высказывать собственное мнение ...

Перевод Ирины Адельгейм

23 процента голосов избирателей, тогда как проельцинские политики — всего 16 процентов.

³³ Александр Солженицын (1918–2008) вернулся в Россию — уже навсегда — в конце мая 1994 года.



ALEKSANDRA SZYMAŃSKA

 ORCID <https://orcid.org/0000-0002-3380-5396>

Uniwersytet Łódzki

Валентина Е. Головчинер, Татьяна Л. Веснина,
Комическое в пьесах М. Булгакова 1920-х годов,
Издательство Томского университета, Томск 2021, 138 с.

Монография *Комическое в пьесах М. Булгакова 1920-х годов* представляет собой комплексное, всестороннее и углубленное исследование драматургии Михаила Булгакова в контексте балаганной традиции. Профессор Валентина Е. Головчинер — специалист в области русской драматургии. Она является автором свыше 300 научных публикаций, среди которых значительное место занимают работы, посвященные проблеме комического в литературе и разным аспектам драматургии. Татьяна Л. Веснина — специалист по драматургии и фельетонистке Булгакова.

Авторы впервые вводят в круг научного изучения ранние фельетоны Булгакова и прослеживают близкие аспекты журналистики и драматургии писателя, утверждая, что, именно, с фельетонов начинается путь Булгакова в большую литературу. Задача, которую ставят перед собой исследователи, заключается в изучении системных проявлений поэтики фельетонов в поэтике драматических произведений писателя. Главным предметом размышлений в монографии стало проявление в этих двух жанрах комического.

Авторы справедливо замечают, что комизм, рядом с раскованной интонацией, широким спектром игрового, свободным выбором тем и объектов, является одной из характерных черт фельетона. Фельетон, как утверждают томские исследователи, трансформировал ряд приемов «балаганной парадигмы».

Открытие Булгакова-фельетониста для широкой читательской аудитории произошло только в пору перестройки. Опубликованные тогда газетные публикации начинающего писателя дали интереснейший материал для размышлений об истоках поэтики и формировании художественной стратегии Булгакова. Ценность исследования Валенти-

ны Головчинер и Татьяны Весниной состоит в том, что они обратились к его ранним фельетонам владикавказского периода, которые до сих пор ускользали внимания исследователей. Итак, авторами анализируются фельетоны *Грядущие перспективы* (1919), *В кафе* (1920) и *Неделя просвещения* (1921). Углубленный анализ названных текстов приводит томских блугаковедов к интереснейшим заключениям. Отмечается полифоничность авторского голосоведения, нарастание компонентов художественной выразительности, соединение лирики и иронии, героического пафоса и сатиры, что станет потом важнейшей чертой драматургии Булгакова. Авторами делается замечание, что одновременно с интересующими их фельетонами возникали первые пьесы Булгакова — *Самооборона* и *Парижские коммунары*, что проливает свет на природу текстопорождения начинающего писателя.

В отдельной главе рассматривается фельетон *Багровый остров* (1924). Исследователями отмечается использование в фельетоне приема авторской игры, комическая выразительность заголовочного комплекса, семантические перевертыши значений географических названий и имен персонажей, присутствие балаганно-теарализованных сцен. Подробный и внимательный анализ текста позволяет авторам заключить, что *Багровый остров* является примером драматизации фельетонной прозы (использование диалогов, масок, самоиронии и проч.).

Интересной с точки зрения заявленной в монографии проблематики является глава, посвященная драме *Багровый остров*. По мнению авторов, факт сохранения названия фельетона в пьесе, написанной в 1928 г. позволяет думать о прямом указании Булгакова на их внутреннюю связь. Итак, предметом интереса в этой главе становится трансформация фельетонных компонентов в действии пьесы. Как утверждают литературоведы, в драме за счет дополнения новыми сюжетными линиями, развивается потенциал балаганной поэтики (прием удвоения, повторы, путаница, игра словами и смыслами). Анализ пьесы позволяет подчеркнуть существенную роль удвоения, которое рассматривается через призму балаганно-фельетонной выразительности. По мнению исследователей, система удвоений создает особые основания для вариативности интерпретации пьесы *Багровый остров*.

В контексте фельетонов Булгакова первой половины 1920-х гг. рассматриваются в монографии тоже пьесы *Зойкина квартира* и *Дни Турбиных*. В первой из них авторами отмечается связь персонажей пьесы с их предшественниками в фельетонах, использование отдельных тем, приема замены/подмены, а также использование в пьесе одного из характерных средств фельетонной выразительности — метонимии. Во второй пьесе отмечается чередование серьезного и несерьезного (например, использование выразительных средств балаганного типа в представлении некомического персонажа), балаганного с остро дра-

RECENZJE

матическим, что, как замечают авторы, характерно для драматургии Булгакова в целом. Как последняя с точки зрения места и функции комических компонентов в поэтике драм Булгакова анализируется пьеса *Бег*. Отмечаются те ее компоненты, которые созданы на основе форм народного театра, а также фельетонное содержание снов.

Стоит отметить, что все аналитические главы монографии отличаются внимательным и тонким подходом к текстам, глубоким знанием поэтики фельетонов и драмы. Соавторы убедительно доказывают свой тезис о фельетонных истоках драматургии Булгакова, раскрывая ее багланную основу.

Монография будет полезна преподавателям и студентам филологических факультетов, а также всем, кто интересуется драматургией и творчеством Михаила Булгакова.



GRAŻYNA MAROSZCZUK

 ORCID: <https://orcid.org/0000-0001-8482-8523>

Uniwersytet Śląski w Katowicach

Gustaw Herling-Grudziński. *Między Wschodem a Zachodem*,
red. Grzegorz Przebinda, Bartosz Gołąbek, Wojciech Gruchała,
Wydawnictwo Humanistyczne Pigionianum,
Karpacka Państwowa Uczelnia w Krośnie, Krosno 2020

Setna rocznica urodzin Gustawa Herlinga-Grudzińskiego zrodziła potrzebę dyskusji naukowych poświęconych wybitnemu pisarzowi i eseiście. W 2020 roku ukazała się pokonferencyjna praca zbiorowa wydana przez Karpacką Państwową Uczelnię im. Stanisława Pigonia w Krośnie. Spójne tematycznie opracowanie naukowe biografii i twórczości prozaika problematyzuje dokonania pisarza w perspektywie formowania współczesnej tożsamości Europy między Wschodem a Zachodem. Najtrwalsza granica „kulturalna kontynentu europejskiego”¹, pisała Maria Janion, stanowi prawdziwe wyzwanie dla myślenia o naszym miejscu w Europie. Refleksja Herlinga na temat naszej tożsamości kulturowej w „czasach marnych” i sposób, w jaki pisarz „tłumaczył ludziom Zachodu, czym był sowiecki komunizm”² łączyły się — o czym dziś nie wolno zapominać — z pojmowaniem pisarstwa jako posłannictwa i wymagały intelektualnego wysiłku, wprzegającego sprzeczności polityczne w służbę myśli i lektury. Zgodzić wypada się ze stanowiskiem autorów publikacji, że rozpoznawalny komentarz Herlinga — pisarza, diarysty i eksperta w zakresie problematyki komunizmu, pozostającego przez dekady w emigracyjnym odosobnieniu, kształtował myślenie polskich odbiorców, dokonujących trudnej samoidentyfikacji kulturowej w świecie politycznych przesileń. Jest bowiem w tej prozie wyczuwalny namysł, respekt i wzmożone skupienie nad losem człowieka doświadczającego mrocznych stron ludzkiej

¹ M. Janion, *Polska między Wschodem i Zachodem*, „Teksty Drugie” 2003, nr 6, s. 138.

² W. Bolecki, *Twórczość Gustawa Herlinga-Grudzińskiego. W setną rocznicę urodzin pisarza*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *Gustaw Herling-Grudziński. Między Wschodem a Zachodem*, red. Wydawnictwo Humanistyczne Pigionianum, Krosno 2020, s. 48.

egzystencji, samotności, cierpienia, okrucieństwa zbrodniczych systemów. Do książki zaproszono wybitnych uczonych z Polski i z zagranicy, badaczy różnych pokoleń, historyków literatury, tłumaczy, współtwórców inicjatyw wydawniczych, przyjaciół i gości neapolitańskiego domu pisarza. Poznajemy zatem prozaika i jego dorobek w przekrojach analityczno-interpretacyjnych syntez i we fragmentach osobistych refleksji wspomnieniowych. Głosy zebrane w księdze pokonferencyjnej są pokłosiem spotkania wybitnych uczonych, którzy zechcieli podzielić się doświadczeniem wielu lat badań na gruncie herlingologii i młodych badaczy, dla których dokonania Herlinga pozostają twórczą inspiracją. W uporządkowaniu kompozycyjnym monografii wypada zaznaczyć tematyczny podział zebranych szkiców na konteksty „wschodnie” i „zachodnie”. Doświadczenie łagrów na Wschodzie Europy i osamotnienie emigranta na Zachodzie — jak powtarzał Herling — tworzą główne tematy jego refleksji twórczej. Część pierwsza *Wschód — opowiedzieć inny świat* stanowi próbę możliwie szerokiego omówienia kontekstów literackich dzieła Herlinga z nawiązaniem do tradycji literatury świadectwa Zagłady. Przypomnieć wypada, że to właśnie w łagrach sowieckich autor *Innego Świata* narodził się jako pisarz. Część druga, zatytułowana *Zachód — emigracja, los*, traktuje o emigracyjnych doświadczeniach Grudzińskiego, mieszkającego i tworzącego do końca życia na obczyźnie. Tu Herling pozostawał wyspą osobną, samotnikiem poszukującym formy, która oddałaby osobisty dramat człowieka, mocującego się z upiorami własnej przeszłości. Jeśli o ramie delimitacyjnej książki mowa warto odnotować trafną decyzję sporządzenia uzupełnień do bibliografii, zamieszczenia archiwalnych fotografii pisarza, fragmentów korespondencji i dokumentacji udostępnionej z zgodą Marty Herling, córki pisarza, odpowiedzialnej za archiwum Gustawa Herlinga-Grudzińskiego w fundacji Biblioteka Benedetto Croce w Neapolu. Książka opatrzona została starannymi przypisami, nie tylko w funkcji skwitowania źródeł cytowanych dzieł z zakresu rozległych badań nad twórczością Herlinga, ale i w konieczny sposób odciążających główny wywód od uzupełnień, które są bardzo istotne, ale mogłyby zakłócić tok rozważań.

Monografię zbiorową otwiera komentarz wprowadzający. Grzegorz Przebinda w artykule zatytułowanym *Drugie życie Gustawa Herlinga Grudzińskiego. Próba świadectwa* pyta o to, kim był Herling-Grudziński dla pokolenia autora w latach osiemdziesiątych, kiedy Polacy w kraju i na emigracji odkrywali zachłannie pisarza dla siebie. Wiemy, że Herling należał do czynnych pisarzy polskiej emigracji, w kraju miał zakaz publikowania. Utwory prozatorskie i publicystyka Herlinga drukowane w wydawnictwach podziemnych, przypomina badacz, pełniły funkcję duchowego przewodnika po historii, polityce Wschodniej i Zachodniej Europy. Studium Włodzimiera Boleckiego *Twórczość Gustawa Herlinga-Grudzińskiego. W setną rocznicę urodzin pisarza* stanowi próbę obrysowania konturów całości dokonania literackich autora *Innego Świata*. Perspektywa ta, ujawniona w postępowaniu badawczym, stanowi jednak zasadniczy horyzont dociekania specy-

fiki twórczości Herlinga, odpowiadającej na wyzwania „zablokowanego doświadczenia własnej tożsamości”³. Herling zmagał się z traumą wspomnień dzieciństwa i pamięcią żydowskiego pochodzenia przodków. Doświadczenie to wyrzucone na „brzeg milczenia”, „zamurowania”, niemożności wypowiadania traumy, wyprzedza doświadczenie łagrów i emigracyjnego osamotnienia autora *Wieży*. Zarazem w obszernych komentarzach, poświęconych biografii i poetyce utworów literackich Herlinga, badacz wyznacza zagadnienia przekrojowe: „poszukiwanie własnej formuły pisarstwa”, „sztukę przedstawiania innego świata”, „tematy *Dziennika pisanego nocą*” i problematykę „dyskursu autobiograficznego”. W toku wywodu dominuje perspektywa porządkująca dorobek literacki Herlinga: „Po pierwsze są to utwory literackie powstałe z przetworzenia fikcji i dokumentu, z doświadczenia osobistego i cudzego, przede wszystkim *Inny Świat*, ale także *Dziennik pisany nocą*. Na pozór są to tylko zapiski diaryistyczne, faktycznie nawet *Dziennik pisany nocą* to bardzo starannie skomponowany utwór literacki. Po drugie, są to opowiadania, z których pierwsze utwory (*Księżę niezłomny*) powstały wkrótce po zamieszkaniu Herlinga w Neapolu. Herling pisał opowiadania przez całe życie — najdosłowniej — aż do śmierci. Pierwsze miały charakter niewielkich opowieści biograficznych o ludziach spotkanych w sowieckim łagrze. Jest ich wiele w *Innym Świecie*. Dwa ostatnie opowiadania zostały odnalezione przez żonę i córkę pisarza już po jego śmierci. Po trzecie — są to eseje poświęcone literaturze — głównie polskiej, rosyjskiej, włoskiej i angielskiej — a także sztuce, religii, i podróżom po Italii. Eseje — będące mistrzowskim połączeniem literackich opisów pejzażu, biografii, refleksji nad kulturą i losem człowieka w historii — stawiają Herlinga wśród europejskich mistrzów tego dyskursu literackiego. Po czwarte wreszcie, jest to niezwykle bogata twórczość publicystyczna, polityczna, filozoficzna, kulturalna i społeczna, której Herling zawdzięczał bardzo dużą popularność na emigracji, a po 1989 roku także w Polsce — gdy już została zniesiona cenzura⁴. Uzupełnieniem syntezy jest wątek żydowskich korzeni przeszłości prozaika. Autor stawia pytanie, czy „żydowski rodowód rodziny Herlinga miał znaczenie w jego późniejszej pisarskiej biografii”⁵. Rekonstrukcja genealogii rodzinnej pisarza znalazła omówienie, powtórzmy za autorem komentarza, w studium *Żydowska rodzina Gustawa Herlinga-Grudzińskiego*, zamieszczonym w piętnastym tomie *Dzieł zebranych*, w dziale *Varia*.

Blok rozważań *Wschód — opowiedzieć inny świat* otwiera Irena Furnal artykułem *Dobryś Bryczkowska i Jacob Josef Herling-Grudziński*. Autorka szkicu przybliży konteksty wczesnego dzieciństwa i młodości pisarza z perspektywy wyznaczonej przez badania topografii „małej ojczyzny”. Studium podejmuje mało znane odbiorcy konteksty losów rodziny i biografii pisarza, jak wątek wiarygodności metryki urodzenia Herlinga-Grudzińskiego, daty

³ Tamże.

⁴ Tamże, s. 26–27.

⁵ Tamże, s. 44.

i okoliczności śmierci Józefa Jakuba Herlinga-Grudzińskiego, losy majątku rodziny pisarza. To oczywiście, że wobec białych plam historii i niekompletnych świadectw rodzinnych, wyjaśnia uczona, nie udało się znaleźć odpowiedzi na wiele pytań. Analiza krytyczna źródeł archiwalnych, epistolografii, różnych wersji informacji pozyskanych bezpośrednio od świadków, stawia nas nad przepaścią suchedniowskiej przeszłości pisarza. Przeszłości, która zniknęła bezpowrotnie w Zagładzie polsko-żydowskich miasteczek. W odnotowanej przez badaczkę wypowiedzi Herlinga z *Dziennika pisanego nocą* czytamy:

Wszystko, co było niegdyś moim dzieciństwem nie pozwalało się wypowiedzieć, bo brakło dla niego słów w ludzkiej mowie i napotykało na suchy gruzełek w krtań, co nie oznacza przecież, że było martwe. Przeciwnie buzowało w podziemiach psychiki jak lawa w głębi wulkanu⁶.

Wspomnieniowa rzeczywistość z jednej strony nieobecna, ukryta, wyparta ujawnia siłę rekonfigurującą myślenie pisarza o sobie. Nielatwe to doświadczenie, stanowiące specyficzny splot tego, co żydowskie i czerpiące z nieżydowskiego otoczenia prozaika. Zarysowana zatem wcale nie w jednoznaczny sposób tożsamość Herlinga dochodzi do głosu w literaturze, jedynym medium ekspresji. Skrupulatny duch rzetelnej lektury, sprawdzający hipotetyczną listę pism kulturalno-literackich i książek młodości autora *Dziennika pisanego nocą*, towarzyszy zestawieniom tekstów kultury Rosji w artykule Feliksa Tomaszewskiego *Długie terminowanie. Gustaw Herling Grudziński i rosyjskość*. Horyzont dociekania specyfiki tych młodzieńczych odczytań, które były lub mogły być punktem odniesienia dla pisarza, opiera się na wieloaspektowym sprawozdaniu historii recepcji twórczości Herlinga. Będą to, powtórzmy za autorem, różne źródła „długiego terminowania” pisarza w szkole wybranych roczników „Wiadomości Literackich” i lektury wyznaczone kanonem edukacyjnym w okresie międzywojennym. Z klaryfikacji podtrzymywanych przez lekturowe wzorce wyrasta idealizowany wizerunek sowieckiej wizji społeczeństwa, konfrontowany z doświadczeniem łagru w Jerciewie:

Pewne jest, że Gustaw Herling-Grudziński dojrzewał w świecie, w którym polskość była jeszcze bardzo młoda, w którym młoda była sowieckość i młody był faszyzm, dojrzewał pod presją programów szkolnych, wydawniczych propozycji, prawniczych i lewicowych czasopism⁷.

Linia postępowania badawczego została bardzo wyraźnie wytyczona już na wstępie artykułu Tadeusza Sucharskiego *Czy Inny Świat jest rzeczywi-*

⁶ G. Herling-Grudziński, *Dziennik pisany nocą 1993–1996. Pisma zebrane*, Z. Kudelski (red.), t. 10, Czytelnik, Warszawa 1998, s. 138.

⁷ F. Tomaszewski, *Długie terminowanie. Gustaw Herling -Grudziński i rosyjskość*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *G. Herling -Grudziński. Między Wschodem a Zachodem...*, s. 116.

ście inny? Dzieło Herlinga na tle pierwszych tekstów lagrowych. Interesująco przedstawia się tu kontekst literatury poprzedzającej publikację *Innego Świata* oraz proces wypracowywania przez Herlinga własnej poetyki dzieła literackiego. Temu wszak poświęcony jest erudycyjny komentarz: zbadaniu punktów wyjścia i punktów dojścia prozatorskiej refleksji Herlinga, prześledzeniu szerokich traktów i wąskich ścieżek, kontekstów szczegółowych literatury lagrowej: polskiej i rosyjskiej. Poznajemy tu Herlinga jako krytycznego czytelnika i recenzenta prozy doświadczenia okupacyjnego. W tej sytuacji istotne wydaje się rozróżnienie Hannah Arendt między poznaniem, będącym rezultatem dążenia do prawdy, a poszukiwaniem sensu, myśleniem o losie człowieka poddanego okrutnemu terrorowi sowieckiemu. Dla pisarza wszechobecne zło i okrucieństwo przygniatające człowieka stało się impulsem do przemyśleń o charakterze etycznym. Grudziński wypracowywał, podkreśla badacz, poetykę dzieła literackiego, i zadanie pisarskie postrzegał w kategoriach misji do ujawnienia zbrodniczego systemu totalitarnego⁸.

Isabella Sariusz-Skąpska w szkicu *Pamięć zakłamana, pamięć odzyskana, pamięć zbanalizowana: na przykładzie Zbrodni Katyńskiej* dokonuje osobistej lektury prozy Herlinga. Interesuje autorkę szkicu konfrontacja szkolnych praktyk czytania *Innego Świata* i rocznicowej publicystyki z lekturą konkretnych świadectw losów ofiar Katynia. Jak pisze Maria Kobielska w książce poświęconej *Polskiej kulturze pamięci w XXI wieku*, pamięć o Katyniu usuwana przez długi czas z powodów politycznych (ograniczenia instytucjonalne, zapisy cenzury) stała się po roku 1989 jednym z centralnych wydarzeń polskiej kultury memorialnej⁹. Równoległe jednak z procesem ujawniania zbrodni na jeńcach wojennych zamordowanych przez NKWD wiosną 1940 roku, pisze Sariusz-Skąpska, następował proces banalizacji pamięci. Szkic podejmuje problem nieuchwytności wydarzenia i zniekształcenia pamięci o nim. Badaczka pisze o własnej drodze docierania do literackich dokumentów „pod prąd totalitaryzmu”, o lekturze konkretnych świadectw wydobytych z „dołów śmierci”, o paryskich kwerendach i mobilizacji wysiłku w poszukiwaniu książek wydawanych w oficynach emigracyjnych, niedostępnych w PRL-u i objętych zapisem cenzury prewencyjnej. Bezsilności wobec banalizacji pamięci i anonimowości „symbolicznych grobów” przeciwstawia Sariusz-Skąpska kontrnarrację, która jest próbą zrekonstruowania ogromu straty: projekty, konkursy, spotkania z bliskimi ofiar: „To jest właściwy trop: losy konkretne, zilustrowane fotografiami jednego człowieka,

⁸ T. Sucharski, *Czy Inny Świat jest rzeczywiście inny?*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *G. Herling -Grudziński. Między Wschodem a Zachodem...*, s. 145.

⁹ M. Kobielska, *Pamięć katyńska-warunki możliwości*, w: tejsze, *Polska kultura pamięci w XXI wieku. Zbrodnia katyńska, powstanie warszawskie i stan wojenny*, Instytut Badań Literackich PAN, Warszawa 2016, s. 43–45.

pokazanego w pracy, czy w gronie rodzinnym — to trafia do wyobraźni i pamięci młodych ludzi¹⁰. Temat podjęty przez Izabellę Sariusz-Skąpską ma walor społeczno-kulturowej aktualności. Przedstawiane rozważania wpisują się w dyskusję na temat praktycznej wartości badań humanistycznych, poświęconych zgadnieniom polskiej pamięci zbiorowej.

Szkic Jędrzeja Piekary *Michał Heller i Gustaw Herling-Grudziński — trzydzieści lat przyjaźni i współpracy* traktuje o relacji pisarza i sowieckiego emigranta, sowietologa, literaturoznawcy Michała Hellera. To historia zapisana fragmentarycznie w zapisach dziennikowych Herlinga i w korespondencji intelektualistów z lat 1968–1997. Tu eksponowanym kontekstem staje się klimat intelektualnego dialogu humanistów związanych z myślą rosyjską i twórczością rosyjskich autorów, dwóch myślicieli, których połączyły doświadczenia z sowieckim komunizmem i ważna dyskusja na temat człowieka sowieckiego (*homo sovieticus*):

Był to jeden z niewielu tematów, które ich podzieliły — Heller stał na stanowisku, że bolszewicy doprowadzili do powstania nowego „gatunku człowieka”; Herling zaś ostro się sprzeciwiał takiemu pogładowi, twierdził, że niemożliwe jest starcie „twardego jądra”, które znajduje się w każdym człowieku, że człowieka nie można aż tak zmienić¹¹.

Optyka przyjęta przez Herlinga zakłada, że w człowieku jest zawsze margines niepodległości, który warunkuje jego autonomię. Interesująco prezentują się artykuły podejmujące problem związków myślowych i intertekstualnych w twórczości Gustawa Herlinga-Grudzińskiego i pisarzy rosyjskich. Wiele tu punktów wspólnych w zaproponowanym sposobie problematyzowania twórczości autora *Innego Świata* i dokonań literackich jego antenatów. Artykuł Grzegorza Przebindy *Nadzieja w Martwym domu. Spotkania Herlinga-Grudzińskiego z Szalamowem* przybliży te aspekty twórczości autora *Opowiadań kołymskich*, które były źródłem inspiracji dla polskiego prozaika. I tu pojawia się rozważany w komentarzu Przebindy problem zachowania człowieka w sytuacjach ekstremalnych i postawa Herlinga — rzecz jasna niosąca sensy — przeciwstawiająca ideologicznej przemocy i okrucieństwu pierwiastek ludzki.

Artykuł Leonida Malcewa *Gustaw Herling-Grudziński o Borysie Pasternaku. Kwestie estetyczno artystyczne* ukierunkowuje lekturę na „motywy pasternakowskie” z wyeksponowaniem trzech kwestii, które łączą twór-

¹⁰ I. Sariusz-Skąpska, *Pamięć zaklamana, pamięć odzyskana, pamięć zbanalizowana: na przykładzie Zbrodni Katyńskiej*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *G. Herling-Grudziński. Między Wschodem a Zachodem...*, s. 153.

¹¹ J. Piekara, *Michał Heller i Gustaw Herling-Grudziński — trzydzieści lat przyjaźni i współpracy*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *G. Herling-Grudziński. Między Wschodem a Zachodem...*, s. 185.

czość Pasternaka i Herlinga. Sedno szkicu stanowią zagadnienia „fenomenu ‘zbiegu okoliczności’”, „istoty snu”, oraz „mitu kulturowego Wenecji”¹².

Tragicznym zrządzeniem losu była dla Gustawa Herlinga-Grudzińskiego emigracja neapolitańska. W Neapolu, mieście znienawidzonym i kochanym, prozaik mieszkał ponad czterdzieści lat. Proces „italianizacji”¹³, spotkania z obcą kulturą dał pisarzowi także impuls twórczy, który urzeczywistnił się w jego kreacjach literackich i zapisach diariuszowych. Szkic Jacka Hajduka *Capri, Tyberiusz i Herling-Grudziński* otwiera drugą część rozpraw naukowych i krytycznych poświęconych emigracyjnemu doświadczeniu pisarza. Artykuł traktuje o antycznych tematach i kontekstach w twórczości autora *Wieży*. Punktem odniesienia analiz wybranych fragmentów *Dziennika pisanego nocą* Gustawa Herlinga-Grudzińskiego jest historia pobytu cesarza Tyberiusza na wyspie Capri, dostępna między innymi w historiograficznych i biograficznych przekazach Korneliusza Tacyta i Gajusza Swetoniusza Tranquillusa. Nadto z perspektywy ciągłości tradycji, historia ta była interpretowana i reinterpretowana w późniejszych analizach brytyjskiego pisarza Normana Douglasa. Analiza wybranych fragmentów *Dziennika pisanego nocą* wykazuje, że namysł pisarza nad zafałszowaną „tajemnicą Tyberiusza” ma charakter alegoryczny:

Na Capri (tej maleńkiej wyspie, tym ‘zamkniętym drobnym odprysku świata’ spotkały się wszystkie ważniejsze prądy myślowe i ideologiczne, kształtujące człowieka współczesnego autorowi *Innego Świata* i *Dziennika pisanego nocą*. Rzymski imperializm, judaizm, chrześcijaństwo, marksizm-leninizm (rewolucja), kapitalizm a wreszcie faszyzm. Capri, wyspa dziejowych wiatrów, to wielka historiozoficzna metafora, a zarazem obok wulkanu jeden z głównych filarów *Dziennika pisanego nocą*¹⁴.

Autor wykazuje we własnej lekturze diariuszowych zapisów Herlinga, że pełniejszy i bogatszy Tyberiusz wylania się z lektury dynamicznej, aktualizującej, stawiającej opór zamkniętym fragmentom przeszłości i czarnej legendy „cesarza dewianta”, który ukrył się na wyspie przed życiem politycznym, oddając uciechom cielesnym. Studium Iriny Adelgejm *Oblicze Boga, oblicze Losu – cień terremonto jako narzędzie przystosowania* traktuje o doświadczeniach neapolitańskich Herlinga związanych ze strefą aktywności sejsmicznej Wezuwiusza w Zatoce Neapolitańskiej. W trzynastym rozdziale *Rozmów w Dargonei*, inspirowanych wywiadem z Włodzimierzem Bolec-

¹² L. Malcew, *Gustaw Herling-Grudziński o Borysie Pasternaku. Kwestie estetyczno-artystyczne*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *Gustaw Herling-Grudziński...*, s. 235.

¹³ M. Śniedziewska, „Osobiste sprawy i tematy”. *Gustaw Herling-Grudziński wobec dwudziestowiecznej literatury włoskiej*, Instytut Badań Literackich PAN, Warszawa 2019, s. 56.

¹⁴ J. Hajduk, *Capri, Tyberiusz i Herling-Grudziński*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *Gustaw Herling-Grudziński...*, s. 253.

kim, Herling zamieścił obszerny komentarz na temat opowiadania *Gruzy*: „literackiego requiem po pierwszym trzęsieniu ziemi”¹⁵ i szoku, jakiego doznaje człowiek wobec zakrytej twarzy żywiołu, z którym nie sposób się skonfrontować. W „dragonejskich” rozmowach z Włodzimierzem Boleckim, Herling, przywołując historię francuskiego wydania *Innego Świata*, nie krył swego rozgoryczenia krzywdzącym potraktowaniem łagrowej epopei. Kontekst ten znajduje znakomite dopowiedzenie w szlucu Marii Dealaperrière *Francja w oczach Herlinga*. Jak pisze badaczka: „Herling powracał z goryczą do dręczących go pytań: czym kierowało się wydawnictwo Plon, rezygnując z wydania książki, mimo uprzednio wyrażonej zgody na jej publikację? [...]”¹⁶. Szkic przynosi próbę możliwie szerokiego spojrzenia na konteksty towarzyszące ponadtrzydziestoletnim niepowodzeniom francuskiego wydania opowieści o sowieckiej cywilizacji więziennej. Interesująco przedstawiają się tu podrozdziały: *W kręgu francuskich intelektualistów*, *Inspiracje pisarskie*. Co istotne, wyjaśnienia *Dramatycznych losów „francuskiego” Innego Świata* chyba zawodzą pozornie czytelnymi kluczami politycznymi czy koninkaturalnymi, bowiem pod nimi, pisze badaczka, kryją się osobiste dramaty pisarza i jego krytyczne spojrzenie na intelektualny Paryż przez pryzmat własnych doświadczeń totalitarnych. Wnikliwy komentarz wyrasta z podszeptów lektury *Dziennika pisanego nocą* i korespondencji Herlinga z Józefem Czapskim¹⁷.

Książka poświęcona dokonaniom twórczym Herlinga-Grudzińskiego publikuje po raz pierwszy w języku polskim tekst Domenica Scarpy *Obóz koncentracyjny i lagier. Levi i Herling, pisarze odpowiedzialności*. W uwagach redakcyjnych zamieszczonych pod tekstem czytamy:

Przedstawiamy tłumaczenie tego tekstu ze względu na jego wagę w recepcji prozy Herlinga-Grudzińskiego we Włoszech i ze względu na jego nieobecność w polskiej debacie o Herlingu. Autor, Domenico Scarpa, dzisiaj, jak mówi, wiele rzeczy napisałby inaczej, wiele by pogłębił, ale wspólnie zdecydowaliśmy, żeby nie zmieniać tekstu z roku 1997, dając tym samym świadectwo pewnego, ówczesnego stanu wiedzy i refleksji¹⁸.

Spotykając prozę Herlinga i Leviego w przekrojowym dla ich twórczości temacie obozowej rzeczywistości, wyłoniono istotne, literackie realizacje głównego tematu, silnie obecnego w twórczości obu pisarzy i wykazano

¹⁵ G. Herling-Grudziński, W. Bolecki, *Rozmowy w Dragonei*, Wydawnictwo Szpak, Warszawa 1997, s. 101–103.

¹⁶ M. Delaperrière, *Francja w oczach Herlinga*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *Gustaw Herling-Grudziński...*, s. 274.

¹⁷ Tamże.

¹⁸ D. Scarpa, *Obóz koncentracyjny i lagier. Levi i Herling, pisarze odpowiedzialności*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *Gustaw Herling-Grudziński ...*, s. 306.

różnice w literackim kształtowaniu przeszłości totalitarnych praktyk. Niezwykle cenne jest tu wyeksponowanie problematyki doświadczeń obozu koncentracyjnego w książkach Leviego *Czy to jest człowiek*, *Rozejm*, *Pogrążeni i ocaleni* i *Innym Świecie* Herlinga Grudzińskiego. Recenzje włoskiego tłumaczenia *Innego Świata* dowodzą, że za sprawą literackiego świadectwa Herlinga do opinii publicznej po wielu latach dociera prawda o sowieckich łagrach. Włoska opinia publiczna nie była przygotowana na konfrontację z okrucieństwem zbrodniczych systemów, jak powie Herling, „bliźniaczych wytworów” cywilizacji więziennej XX wieku. Ciekawie przedstawia się tu analiza obu utworów wiodąca do wykazania, tak istotnej dla założeń interpretacyjnych Domenico Scarpa analogii między opisywanymi zjawiskami. Szkic odsłania wiele przekrojów, poprzez które dokonuje się w interpretacji ogląd rzeczywistości więziennej opisaną przez Herlinga i Leviego. Niemierzalne w swym okrucieństwie koncentracyjne uniwersum działa w oparciu o stałe antropologiczne reżimu takie, jak „mentalność poszlakowa”, brak solidarności pomiędzy więźniami, głód, charakterystyczne dla lagerszprachy słowa-klucze, wreszcie biurokracizm. Niemiecki, precyzyjny, „taylorowski urzędnik” i „stalinowski misjonarz rewolucji” decydują w bezdusznej rutynie koncentracyjnej zony, o losie ludzi skazanych na Zagładę.

Drugą część rozważań, poświęconych emigracyjnej biografii pisarza zamykają artykuły, które dokumentują spotkania, rozmowy z pisarzem, czyniąc głównym źródłem relacji pokłady wdzięcznej pamięci. Komentarz wspomnieniowy Romana Sosnowskiego *Moje spotkania z Gustawem Herlingiem-Grudzińskim w Neapolu w 1992 roku* traktuje o wizycie autora w neapolitańskim domu pisarza. Interesuje autora trudna, skomplikowana sytuacja emigranta, jego samotność w środowisku elit intelektualnych Włoch. Herling zgromadził imponujący, publicystyczny dorobek.

Teraz, kiedy wiem, jak wiele napisał i opublikował we Włoszech, mam więcej wątpliwości. Niewątpliwie natomiast jest wspomnienie jego gorzkich słów o elitach włoskich i przekonanie, które wyraził w mojej obecności [...], że we Włoszech nie został odpowiednio doceniony¹⁹.

Wraz ze zmianami ustrojowymi w Polsce, zmieniło się nastawienie elit włoskich do publicystyki Herlinga. Wizyta pisarza w Polsce w 1991 roku wywołała uzasadnione zainteresowanie osobą twórcy słabo znanego polskiemu odbiorcy. W tym czasie powstają profesorskie książki-wywiady Włodzimierza Boleckiego z Herlingiem-Grudzińskim (*Rozmowy w Dargonei*, *Rozmowy w Neapolu*). Pozycję osobną zajmuje *Widok z wieży*, publikacja Elżbiety Sawickiej złożona z kilku wywiadów publikowanych wcześniej w „Plus Minus” i „Przeglądzie Polskim”. Przykład książki z wycuciem do-

¹⁹ R. Sosnowski, *Moje spotkanie z Gustawem Herlingiem -Grudzińskim w Neapolu w 1992 roku*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *Gustaw Herling-Grudziński...*, s. 313–314.

bręgo dziennikarstwa, zakreślającej wizerunek pisarza, publicysty, diagnosty spraw publicznych. Autorka, odnosząc się do różnych materiałów (fragmenty korespondencji, teksty z archiwum, fotografie), koncentruje się na ważnych wydarzeniach z życia pisarza, jego związkach i rozstaniu ze środowiskiem „Kultury” paryskiej (*Mój wielki szef*). Kontekst ten przybliży Sawicka w artykule *Po rozstaniu z „Kulturą” paryską. Ostatnie lata Gustawa Herlinga-Grudzińskiego*, który jest próbą rekonstrukcji tamtych wydarzeń. W szkicu przywołano historię listu z 1996 roku, w którym Herling decyduje się publikować fragmenty dziennika w tygodniowym dodatku do „Rzeczpospolitej” — „Plus Minus”, rozpoczynając pięcioletnią współpracę z redakcją prestiżowego magazynu publicystycznego. Artykuł sprawozdaje aktywność i działania pisma, publikacje w wydawnictwie „Presspublica”, promocje nowych opowiadań Herlinga, relacje z dyskusji wokół laudacji i doktoratów honorowych, które pisarz otrzymał w latach dziewięćdziesiątych. By pozostać w kręgu tematów poruszanych w szkicu, wskazać trzeba, powracający w różnych odsłonach komentarz do tematów wcale niełatwych takich, jak: kryzys przyjaźni wybitnych intelektualistów Herlinga i Jerzego Giedroycia, wątku żydowskiego w biografii pisarza i senilnej refleksji literackiej prozaika w ostatnich, autobiograficznych opowiadaniach. Książkę zamyka komentarz bibliograficzny: *Uzupełnienie do bibliografii Gustawa Herlinga-Grudzińskiego* Janusza S. Gruchały z tekstem nieznannej bibliografom przedmowy, poprzedzającej, jak pisze autor, francuskojęzyczne wydanie opowiadań Józefa Hena, ukrywającego się pod pseudonimem Korab. Szkic podejmuje próbę zrekonstruowania historii anonimowej przedmowy do francuskiego przekładu *Oka Dajana*. Precyzyjny wywód na temat dokumentacji odnalezionej w archiwach pomyślany został jako uzupełnienie emigracyjnej biografii pisarza.

Książka wydana przez Wydawnictwo Humanistyczne Pigionianum stanowi interesujące, przenikliwe analitycznie omówienie spuścizny pisarza znanego w obiegu międzynarodowym:

Twórczość Herlinga zapewniała związek z najwyższymi nurtami myśli i sztuki europejskiej. Równocześnie dzięki tłumaczeniom, czytelnikom w Europie Zachodniej przynosiła wiedzę o Europie Wschodniej — o tej „innej Europie”, oddanej we władanie Stalinowi na konferencjach w Teheranie (1943) i w Jałcie (1945)²⁰.

Wybór tematu jest gestem tym bardziej istotnym, że wpisuje przedstawione rozważania w toczącą się dyskusję na temat aktualnych spraw o pierwszoplanowym znaczeniu historycznym, społecznym i kulturowym. Analiza zebranych w monografii zbiorowej wypowiedzi pozwala stwierdzić, że różnie profilowane wykładnie dokonań pisarskich Herlinga, cenione-

²⁰ W. Bolecki, *Twórczość Gustawa Herlinga-Grudzińskiego. W setną rocznicę urodzin pisarza*, w: G. Przebinda, B. Gołąbek, W. Gruchała (red.), *Gustaw Herling-Grudziński...*, s. 24.

RECENZJE

go komentatora życia politycznego i intelektualnego w Europie, ujawniają także nieoczywistość jego lektury na przykład w kontekście domniemanego stosunku Herlinga wobec żydowskiego rodowodu swojej rodziny. Widziany z tej perspektywy Gustaw Herling-Grudziński pozostaje nieznanym pisarzem również dzisiaj.



ЕЛЕНА ВАЖЕНИНА

ORCID: <http://orcid.org/0000-0003-0600-1242>

Донецкий национальный университет имени Василя Стуса

АНАТОЛИЙ ЗАГНИТКО

<http://orcid.org/0000-0001-7398-6091>

Донецкий национальный университет имени Василя Стуса

РУССКАЯ ФРАЗЕОЛОГИЯ КАК СУДЬБА

Анатолий Жуков, *Очерки по фразеологической семантике: монография, 2-е издание, переработанное и дополненное*,
Издательство РУСАЙНС, Москва 2020

Говорить о моде в науке не принято, однако это не означает, что таковой не существует. Современные тренды в науке распространились и на лингвистику, что обусловило появление ряда новых «наук» на стыке двух, а то и трех направлений — этнолингвистика, психоллингвистика, социоллингвистика, компьютерная лингвистика и т.д. Все меньше становится исследований по диалектологии и фразеологии, в которых никак не обойтись без огромного эмпирического материала, его скрупулезного осмысления. Поэтому появление рецензируемой монографии представляется по истине знаковым событием в мире языкознания.

Имя и деятельность Анатолия Жукова связаны с новгородской фразеологической школой, основателем которой является его отец — профессор Влас Жуков. Отрадно, что ученый предстает достойным продолжателем дела своего отца, чье имя стоит в одном ряду с именами самых известных лингвистов XX века — Виктора Виноградова, Юрия Караулова, Соломона Кацнельсона, Елены Кубряковой, Виктора Мокиенко, Александра Молоткова, Вероники Телии и многих других.

Рецензируемым *Очеркам...* предшествовала кропотливая и последовательная работа автора, что подтверждает целая серия публикаций¹.

¹ А. Жуков, *Фразеологические фантомы*, «Проблемы истории, филологии, культуры» 2009, № 4 (26), с. 128–133; А. Жуков, *О мотивировке и выво-*

Фразеология — яркий лингвистический феномен и одно из наиболее сложных для изучения языковых явлений, не нашедших однозначной интерпретации в лингвистике. В свое время Отто Есперсен назвал фразеологию «деспотически капризной и неуловимой вещью» с яркой выразительностью, образностью и эмоциональностью.

Анализируемая монография состоит из *Предисловий* (с. 5–8), Главы I. *Общие вопросы фразеологической семантики* (с. 9–85), Главы II. *Семантика фразеологической переходности* (с. 86–168), Главы III. *Фразеологическое значение в словаре и в тексте* (с. 169–234), Главы IV. *Русская фразеология как судьба* (с. 235–262), *Заключения* (с. 263–264), *Списка литературы* (с. 265–276), включающего более 200 позиций, *Словарей и справочников* (40 единиц) (с. 276–278).

В первой главе ученый подробно и скрупулезно анализирует проблему уровневой квалификации фразеологических единиц, обращая внимание на их парадигматические и синтагматические связи. Автор основательно излагает противоположные взгляды исследователей на анализируемую проблему (и тех, кто выделяет отдельный фразеологический уровень, пласт, подсистему, и тех, кто считает, что фразеологизмы являются заменителями слов, своего рода языковыми излишками, и тех, кто причисляет фразеологизмы к промежуточным явлениям), заочно дискутируя со сторонниками концепций и аргументируя собственный подход к квалификации фразеологизмов.

Самостоятельный интерес для автора представляют иерархические отношения фразеологизмов:

Присущи ли иерархические отношения фразеологизму? Думается, что нет. Казалось бы, фразеологизм состоит из компонентов, строится из них. Однако аналогия между фразеологизмом и основными единицами языка является ложной. Дело в том, что компонент лишен той данности, качественной определенности, какими наделены фонемы, морфемы или слова. Компонент не существует до фразеологизма, он вторичен, произволен и в этом смысле является системоприобретенным продуктом. Следовательно, внесистемные значения субъективно присваиваются или навязываются компонентам уже после того, как сформировалась фразеологическая единица, и по этой причине не обладают прогнозирующей силой. Они носят потенциальный (вероятностный) характер. Думается, что внесистемных значений нет и не может быть у компонентов до образования фразеологизма².

Интересными являются и аргументы автора в пользу положения о том, что иерархические отношения отсутствуют как внутри фразео-

димости фразеологизмов, «Мир русского слова» 2013, № 2, с. 9–133 и т.д. (всего более 10 единоличных и в соавторстве работ).

² А. Жуков, *Очерки по фразеологической семантике*, 2-е изд., перераб. и доп. Издательство РУСАНАЙС, Москва 2020, с. 10. Дальше указаны страницы этого издания в тексте.

логизма, так и за его пределами. Анатолий Жуков солидарен с Владимиром Архангельским в том, что «над фразеологическими единицами нет других уровней и единиц, по отношению к которым как единицам высшего порядка инварианты единиц фразеологического уровня являлись бы единицами низшего порядка»³. Такое положение фразеологизмов в системе языка ученый объясняет тем, что при целостности значения они имеют расчлененную структуру, а само фразеологическое словосочетание, как и его отдельные части, не только преобразуются семантически, но и утрачивают в той или иной степени свои формальные качества.

Примечательно, что Анатолий Жуков дает определение фразеологизма, которое несколько разочаровывает читателя своей традиционностью. Однако уже в выводах к первому параграфу автор квалифицирует фразеологизм как «гибридную единицу языка, функционально и семантически тяготеющую к слову, а в формальном отношении в силу своеобразной уровневой инерции напоминающую синтаксические единицы» (с. 13).

Несколько непривычно и даже провокационно звучит название параграфа *О природе фразеологического знака*. Логично встраивается тезис Власа Жукова о том, что сущность фразеологического знака обусловлена характером целостности фразеологической единицы⁴, в концепцию Анатолия Жукова о фразеологизме как элементе знаковой системы языка. С этим фрагментом концепции мы не можем согласиться без некоторых возражений. Да, исключить фразеологизм из системы языка не представляется возможным, но и квалифицировать его как знак не имеет смысла. Фразеологизмы не совсем однородны в своем массиве, включая идиомы, словосочетания и под. Первые всегда однозначны, а другие — многозначны, что свидетельствует о различии их природы. Ученые пробуют найти подступы к неприступной внутренней форме идиом, но попытки в одном массиве объединить образования с разной степенью идиоматичности приводят к тому, что в одном кластере оказываются и символы, и знаки. К тому же, словосочетание «символический знак» является языковедческим оксюмороном, поскольку символ и знак противоположны по своей сути.

Далее автор *Очерков...* обращается к вопросу о природе компонентов фразеологических единств и сращений. Долгая история изучения фразеологизмов свидетельствует о том, что вопрос не может быть решен, если оставаться только в пределах отдельных разделов языкознания,

³ В. Архангельский, *Устойчивые фразы в современном русском языке // Основы теории устойчивых фраз и проблемы общей фразеологии*, Издательство Ростовского университета, Ростов на Дону 1964, с. 98–99.

⁴ В. Жуков, *Семантика фразеологических оборотов*, изд. 2, доп., Издательство URSS, Москва 2019, с. 25.

поскольку все существующие концепции можно условно разместить в диапазоне между двумя полюсами: компоненты фразеологизмов считать словами — компоненты фразеологизмов не являются словами. Даже попытка в очередной раз указать на то, что «компонент фразеологизма — явление не одномерное, а многомерное» (с. 18), не меняет ситуацию в корне. Очевидно, для решения данной проблемы целесообразно выйти в сферу лингвокультурологии или этнолингвистики. Анатолий Жуков, пытаясь разорвать замкнутый круг обозначенной проблемы, предлагает обратиться к понятию семантической редукции слова и, уже опираясь на него, выстраивает типологию семантически маркированных компонентов, принимая за ее основу принцип семантического поля.

Автор останавливается на различных типах маркированности компонентов фразеологизмов: 1) лексической, восходящей к полнозначительным словам, доля участия которых в становлении фразеологического значения очевидна; 2) семной, где показателем является только часть исходного содержания слова; 3) символической, когда исходное символическое значение частично удерживаемое в семантике фразеологизма и достаточно регулярно воспроизводимое в ряде фразеологических единиц; 4) коннотативной; 5) архаической; 6) реликтовой (два последних типа представляют собой соединение коннотативной маркированности лексической, прономинальной, категориальной); 7) нумеральной; 8) прономинальной; 9) грамматической; 10) препозиционной; 11) партикулярной; 12) конъюнктивной; 13) деривационной.

Ученый не претендует на описание всех видов маркированности по всем возможным критериям, в чем искренне сознается, но предложенная классификация оказалась объемной, логичной и точной. В результате работы над классификацией Анатолий Жуков приходит к выводу, что «в зависимости от типа семантической маркированности компонент то сближается со словом, то удаляется от него» (с. 25). Семантическая маркированность компонентов понижает идиоматичность фразеологизма. Именно на почве семантической проницаемости компонентов и взаимодействия фразеологизмов с уровневыми единицами языка формируются различные виды фразеологической переходности.

Рассматривая вопрос о существенных свойствах фразеологизмов и их семантической классификации, автор замечает, что любая классификация является результатом теоретических предпочтений и выбора ученого, с чем нельзя не согласиться. Исследователь сознательно выбирает методику фразеологической аппликации Власа Жукова, по которой классифицирующими признаками исследуемых единиц являются идиоматичность, мотивированность, налагаемость и выводимость.

Хотелось бы обратить особое внимание на небольшой по объему, но огромный по своему потенциалу для лингвистов, пара-

граф о фразеологических фантомах в русском языке (с. 60). Судя по предвещающим монографии публикациям, проблема является острой и достаточно интересной для всего научного сообщества.

Автор снова и снова возвращает нас к проблеме разграничения символа и знака в царстве фразеологизмов. Говоря об их категориальном синкретизме, ученый отмечает, что

фразеологический знак вторичен и более сложен в содержательном и формальном плане по сравнению со словесным знаком», т.е. имплицитно исследователь вводит понятие фразеологического знака (отличного от словесного!), главное противоречие которого кроется в соотношении его актуального и этимологического содержания, с одной стороны, и его актуальной и этимологической формы — с другой (с. 68).

Но, к сожалению, далее вывода о том, что фразеологический знак противоречив не только в целом, автор не идет.

Интереснейшим представляется рассмотрение фразеологического потенциала исследуемых единиц, опять же с использованием полевой теории. Анатолий Жуков, задействуя огромный пласт русской фразеологии, находит все новые подтверждения выдвинутой Виктором Виноградовым гипотезе о том, что

[...] чем уже круг синтаксических связей слов, чем ограниченнее его грамматические видоизменения, чем неразложимее его морфологический состав, тем синкретичнее его природа, тем неразрывнее в нем связь лексических и грамматических значений⁵.

Вторая глава посвящена описанию семантики фразеологической переходности с констатацией того, что идеи переходности, отражая объективную реальность, постепенно обобщаются в истории науки в важнейший гносеологический принцип. Подобные научные гипотезы подтверждаются и авторскими исследованиями. Именно переходность помогает ученому разобраться с вопросом о классификации фразеологических единиц, поскольку не существует идеальных, образцовых языковых явлений — у каждого слова своя история, а, следовательно, любая классификация является абсолютно условной. Кроме того, существует и масса фразеологических единиц, которые трудно втиснуть в любые классификационные рамки. Все эти явления Анатолий Жуков рассматривает в рамках фразеологической переходности. В данном ракурсе подробно рассмотрены фразеологизмы с лексически связанным значением, фразеологизмы с общим аналитическим значением, такие явления, как вариантность фразеологизма, лексико-фразеологиче-

⁵ В. Виноградов, *Русский язык. Грамматическое учение о слове*, изд. 2-е, Высшая школа, Москва 1972, с. 34.

ская синонимия в системе межуровневых отношений и слова-дублеты, дефразеологизация и т.д., обладающие высокой познавательной ценностью для любого лингвиста. Не оминут ученый и безэквивалентные фразеологизмы, являющиеся важнейшими показателями номинативно-познавательных и системных свойств устойчивых оборотов (с. 141). Однако данный аспект, по мнению автора, нуждается в более глубоком изучении, а единицы, подпадающие под обозначенные параметры, — в более тщательном описании и систематизации.

Особое внимание в этой главе уделено типологии пословично-поговорочных выражений (с. 145). Являясь сторонниками узкого подхода в понимании фразеологии и используя метод фразеологической аппликации Власа Жукова, авторы (а глава написана в соавторстве с Константином Жуковым) предлагают обновленный вариант классификации паремиологического материала (с. 147). Огромным плюсом является и то, что не оставлено без внимания и вопросы методологии современных паремиологических исследований, предлагая в их изучении опираться на теорию структурно-семантического поля и полагая, что это

в конечном итоге позволит не только гармонизировать виртуальную картину соотношения основных и промежуточных единиц (например, дополнить традиционно выделяемые разряды пословиц, поговорок и пословично-поговорочных выражений промежуточными типами пословичных и поговорочных выражений), но и реально приблизить исследователя к самому объекту (с. 150).

Актуализированы и новые методы типа концептологического, контекстного, фреймового анализа пословично-поговорочных изречений.

Весьма важным является исследование слов-сопроводителей и слов-спутников, представляющих обязательное и необязательное окружение фразеологизма, а также — смежных с фразеологизмами образований (составные производные номенклатурного, терминологического и перифрастического характера, а также отдельные единицы в области русской паремиологии).

Анализ фразеологического значения в словаре и тексте» осуществлен с использованием теории языковой личности, что позволило исследовать и семантическое пространство слова и фразеологизма и использование последнего в различных стилях языка. Ученый небезосновательно утверждает, что вероятностные границы семантического потенциала фразеологизма зависят не только от его парадигматических, синтагматических и деривационных свойств, но и от так называемого «четвертого измерения» — текстового параметра, «ибо семантическое пространство языковой единицы определяется не только системными, но и ее системоприобретенными свойствами» (с. 184).

Автономным представляется рассмотрение проспекта словаря *Фразеологизмы на газетной полосе* (в соавторстве с Мариной Жуковой), основу которого составили наиболее употребительные в текстах современных российских газет устойчивые обороты. Своеобразие лексикографической работы состоит в том, что фактический материал представляют тексты из наиболее популярных газетных изданий 2006–2007 годов. В словаре рассмотрены семантические, стилистические, грамматические и другие свойства фразеологизмов, а в отдельных случаях приводятся историко-лингвистические, исторические, культурологические и этимологические комментарии к ним.

Довольно нестандартной для отечественной фразеологии является тема заимствований из других языков. Примечательно, что автор не просто исследует факты языковых заимствований в русском языке, но и параллельно изучает заимствование фразеологизмов как освоение концептов другой картины мира и адаптацию «импортного продукта» на отечественной почве.

Говоря о современной русской лексикографии, Анатолий Жуков не только восхищается исследованиями отечественных фразеологов Рудольфа Яранцева, Александра Федорова, Александра Молоткова, Власа Жукова, Виктора Мокиенко, но и настаивает на необходимости противостояния плагиату в сфере публикации современных словарей, небезосновательно критикуя издательства за непрофессиональное отношение к делу.

Трогательной является глава *Русская фразеология как судьба*, посвященная памяти учителей — выдающихся фразеологов — Власа Жукова, Николая Лаврова, Виктора Бондаренко, Анатолия Зеленова, — личностей, которые, пройдя военное лихолетье, пришли в науку для неординарных исследований. Выдающиеся ученые — представители новгородской фразеологической школы — оставили после себя не только высочайшего научного качества изыскания, но и глубоко лиричные стихотворные произведения, часть которых посвящена богатству фразеологии родного языка.

В выводах четко сформулировано авторское понимание задач современных фразеологических разработок. Ученый считает, что этот феномен русской речи, по-прежнему представляет собой мало исследованный островок русской лексики, справедливо подчеркивая, что попытки выявить правила сложения смыслов слов в смысл фразеологизма так и остаются всего лишь попытками, а значение фразеологизма, закономерности внутренней комбинаторики его компонентов так и остаются загадкой. Здесь уместно вспомнить философское размышление Аристотеля о том, что сумма частей всегда будет меньше целого. Благодаря трудам выдающихся российских лингвистов, в том числе и автора монографии, в значительной степени удалось раскрыть многие грани фразеологического содержания, его пространство

и структуру, семантические и экспрессивные потенции в тексте и в речевом употреблении.

Текст исследования насыщен иллюстративным материалом, убедительными примерами, корректным использованием исследований предшественников, глубокими, содержательными и профессиональными комментариями, теоретическими обобщениями, любопытными умозаключениями. Язык монографии прост и доступен, изложение логично, последовательно и убедительно. Отрадно, что, продолжая дело своего отца, Анатолий Жуков подарил знатокам и почитателям русской фразеологии еще одну замечательную книгу, содержащую теоретические открытия, воодушевляющую на дальнейшее развитие теории фразеологии и имеющую большой прагматический потенциал.



NOTY O AUTORACH

EWA BIAŁEK

Dr hab., adiunkt w Katedrze Językoznawstwa Słowiańskiego Instytutu Neofilologii Uniwersytetu Marii Curie-Skłodowskiej w Lublinie. Zainteresowania naukowe: teoria, praktyka i dydaktyka przekładu, leksykografia przekładowa, dyskurs dyplomatyczny, etnolingwistyka, współczesny język rosyjski i in. Autorka kilkudziesięciu artykułów naukowych, a także monografii *Kolokacja w przekładzie. Studium rosyjsko-polskie* (2009) oraz dwóch słowników: *Kolokacja w przekładzie. Słownik rosyjsko-polski* (2011), *Dyplomacja i polityka. Rosyjsko-polski słownik przekładowy, A-II* (2019).

Kontakt: e.bialek@poczta.umcs.lublin.pl

DANIEL BORYSOWSKI

Dr, adiunkt w Katedrze Języków Słowiańskich Instytutu Językoznawstwa na Wydziale Filologicznym Uniwersytetu Opolskiego. Tekstolog i frazeograf, autor lub współautor kilku działów *Podręcznego idiomatykonu polsko-rosyjskiego* oraz współautor *Polsko-rosyjskiego słownika par przekładowych*. Autor monografii *Teksty zespolone a problemy leksykografii przekładowej*. Od kilku lat zajmuje się zagadnieniami z obszaru korpusologii — przede wszystkim kwestią gromadzenia i przetwarzania danych tekstowych, modelowania języka na potrzeby systemów automatycznego rozpoznawania mowy.

Kontakt: dborysowski@uni.opole.pl

ANDRZEJ CHARCIAREK

Dr hab., prof. UŚ, slawista, językoznawca, pracuje w Instytucie Językoznawstwa Uniwersytetu Śląskiego w Katowicach. Autor monografii poświęconej polskiem, czeskim i rosyjskim metatekstowym operatorom fatycznym oraz słownika polsko-rosyjskiego [i] rosyjsko-polskiego. Redaktor i współredaktor 12 monografii wieloautorskich. Opublikował ponad 90 prac naukowych z zakresu pragmatyki językoznawczej, językoznawstwa korpusowego, lingwistyki tekstu i leksykografii przekładowej w Polsce, Czechach, Niemczech, Grecji, Francji, Rosji, Korei Południowej, Estonii i na Słowacji. Członek rad naukowych czasopism: „Folia Linguistica Rossica” (Uniwersytet Łódzki), „Philologia Rossica” (Uniwersytet Hradec Králové), „Opera Slavica” (Uniwersytet Masaryka w Brnie) i „Jazyk a kultura” (Uniwersytet Preszowski).

Kontakt: andrzej.charciarek@us.edu.pl

DARIA DAYTER

Adiunkt (Assistant Professor) w Instytucie Języka Angielskiego i Językoznawstwa na Uniwersytecie w Tampere w Finlandii. Dyplom ze specjalnością filologia angielska i translatoryka uzyskała na Rosyjskiej Akademii Chrześcijańskiej Nauk Humanistycznych (Russian Christian Academy for the Humanities) w Sankt Petersburgu. Pracowała jako tłumaczka, a następnie uzyskała stypendium DAAD, które umożliwiło jej ukończenie studiów doktoranckich oraz obronę pracy doktorskiej na Uniwersytecie w Bayreuth w Niemczech w 2014 roku. Stopień doktora habilitowanego uzyskała na Uniwersytecie w Bazylei w Szwajcarii na podstawie prac poświęconym badaniom nad zróżnicowaniem języka w tłumaczeniu ustnym. Jej książka *Discursive Self in Microblogging; Speech Acts and Self-praise* ukazała się w 2016 r. nakładem wydawnictwa John Benjamins.

Kontakt: daria.dayter@tuni.fi

ŁUKASZ GRABOWSKI

Dr hab., prof. UO, zatrudniony w Instytucie Językoznawstwa Uniwersytetu Opolskiego. Jego zainteresowania naukowe dotyczą językoznawstwa korpusowego, formuliczności języka naturalnego, przekładoznawstwa oraz komputerowych metod analizy tekstu. W 2013 roku uzyskał stypendium postdoktorskie w Centrum Zaawansowanych Badań nad Językiem Angielskim (CARE) na Uniwersytecie w Birmingham w Wielkiej Brytanii. Pracował również na Uniwersytecie w Ostrawie w Czechach, a także w projektach z zakresu językoznawstwa korpusowego na Uniwersytecie Łódzkim oraz na Uniwersytecie Aston w Birmingham. Prowadził badania z zakresu stylistyki korpusowej i przekładoznawstwa, a także zróżnicowania językowego w obrębie angielskiego dyskursu farmaceutycznego. Obecnie interesuje się wpływem czynników kontekstowych na formuliczność tekstów pisanych i mówionych.

Kontakt: lukasz@uni.opole.pl

ALEKSANDER KIKLEWICZ

Prof. dr hab., pracuje w Instytucie Dziennikarstwa i Komunikacji Społecznej Uniwersytetu Warmińsko-Mazurskiego w Olsztynie. Kierownik Zakładu Komunikacji Społecznej i Języka Mediów. Dyrektor Centrum Badań Europy Wschodniej UWM w Olsztynie. Redaktor naczelny czasopisma „Przegląd Wschodnioeuropejski”. Redaktor trzech serii wydawniczych. Autor ponad 450 publikacji, w tym 21 monografii. Publikował w wielu językach i w wielu krajach. Przedmiotem zainteresowań jest teoria i filozofia języka, metodologia badań lingwistycznych, socjologia językoznawstwa, gramatyka funkcjonalna, składnia semantyczna, komunikologia i pragmatyka językowa, psycholingwistyka i semantyka kognitywna, semantyka logiczna, językoznawstwo konfrontatywne: polsko-wschodniosłowiańskie oraz słowiańsko-germańskie. Kierownik bądź wykonawca kilku projektów naukowo-badawczych, w tym realizowanych we współpracy międzynarodowej.

NOTY O AUTORACH

Jest członkiem Komisji Stylistycznej oraz Komisji Lingwistyki Dyskursu przy Międzynarodowym Komitecie Sławistów, a także członkiem Societas Humboldtiana Polonorum. Jest (powtórnie) członkiem Zarządu Polskiego Towarzystwa Językoznawczego.

Kontakt: akiklewicz@gmail.com

MACIEJ LABOCHA

Dr, absolwent języków wschodniosłowiańskich na Uniwersytecie Śląskim. Ukończył studia podyplomowe na kierunku handel zagraniczny na Akademii Ekonomicznej im. K. Adamieckiego w Katowicach oraz uzyskał certyfikat tłumacza simultanicznego w Centrum Szkoleń Tłumaczeniowych w Sosnowcu. Jest autorem publikacji głównie z zakresu leksykografii przekładowej.

Kontakt: maciej.labocha@wp.pl

YEVHEN LIASHCHEVSKYI

Doktorant w Szkole Doktorskiej Nauk Humanistycznych Uniwersytetu Jagiellońskiego, absolwent filologii rosyjskiej na Uniwersytecie Jagiellońskim w Krakowie, literaturoznawca. Zainteresowania naukowe: teoria i historia literatury, reprezentacja literacka, proza XIX–XXI wieku, literatura obcojęzyczna wydawana w Rosji.

Kontakt: yevhen.liashchevskyi@doctoral.uj.edu.pl

GRAŻYNA MAROSZCZUK

Dr hab., profesor Uniwersytetu Śląskiego w Katowicach, historyk literatury polskiej. Jej zainteresowania badawcze obejmują; prozę polską XX wieku, zagadnienia prozy autobiograficznej i interlokucyjnej. Autorka monografii i artykułów poświęconych między innymi prozie Andrzeja Szczypiorskiego, Tadeusza Konwickiego, wywiadom książkowym Gustawa Herlinga-Grudzińskiego, Juliana Strykowskiemu.

Kontakt: grazyna.maroszczuk@us.edu.pl

MARIA MOCARZ-KLEINDIENST

Dr hab., prof. KUL, kierownik Katedry Translatoryki i Języków Słowiańskich w Instytucie Językoznawstwa Katolickiego Uniwersytetu Lubelskiego Jana Pawła II, redaktor naczelna czasopisma „Roczniki Humanistyczne” z. Glottodydaktyka, redaktor działu Przekładoznawstwo „Półrocznika Językowego Tertium”. Zainteresowania naukowe: translatoryka audiowizualna, interkulturowość w przekładzie tekstów użytkowych, leksykografia dwujęzyczna, przekład a gatunek filmowy, audiodeskrypcja w filmie.

Kontakt: maria.mocarz-kleindienst@kul.pl

GRZEGORZ PRZEBINDA

Filolog rusycysta i historyk idei, wieloletni dyrektor instytutu Filologii Wschodniosłowiańskiej UJ, od 2005 kierownik Katedry Kultury Słowian

Wschodnich UJ. W latach 2012–2020 — rektor Karpackiej Państwowej Uczelni im. Stanisława Pigoń w Krośnie. Redaktor naczelny rocznika „Studia Pigońiana”. Autor ponad 350 publikacji, w tym dziewięciu książek na temat kultury i historii Rosji, Ukrainy i Białorusi. W 1991 uhonorowany Nagrodą Polcul za „współpracę polsko-rosyjską w walce o demokrację”. Ostatnia książka — *Между Краковом, Римом и Москвой. Русская идея в новой Польше* (Moskwa 2013). W 2016 wraz z żoną Leokadią Anną i synem Igozem opublikował przekład *Mistrza i Małgorzaty* z filologicznym komentarzem. Odznaczony Krzyżem Kawalerskim Orderu Odrodzenia Polski (2015) oraz Medalem Gloria Artis (2018). Ostatnio współredagował — wraz z Bartoszem Gołąbką i Wojciechem Gruchałą — zbiorową monografię *Gustaw Herling-Grudziński. Między Wschodem a Zachodem* (Krosno 2020).

Kontakt: gregp6@gmail.com

ANDRZEJ ROMANOWSKI

Prof. dr hab., wieloletni kierownik Katedry Kultury Literackiej Pogranicza na Wydziale Polonistyki UJ, redaktor naczelny Polskiego Słownika Biograficznego, autor wielu książek, ostatnio wydał *Pamięć gromadzi prochy* (Kraków 2021).

Kontakt: arom44@wp.pl

ROMAN ROSZKO

Dr hab., prof. Instytutu Sławistyki PAN. Absolwent Uniwersytetu Warszawskiego. Autor lub współautor pięciu monografii, blisko stu trzydziestu artykułów naukowych; redaktor lub współredaktor ponad dwudziestu monografii wieloautorskich. Sławista, bułgarysta, lituanista, rusycysta. Zajmuje się językoznawstwem synchronicznym, zwolennik teoretycznych badań kontrastywnych z semantycznym językiem pośrednikiem. Od dwudziestu lat pasjonuje się lingwistyką korpusową i komputerowym przetwarzaniem języka naturalnego. Jest członkiem Rady Konsorcjum CLARIN-PL, autorem lub współautorem piętnastu korpusów jedno-, dwu- i wielojęzycznych. Zajmował się również kulturą i językiem mongolskim oraz dialektologią. Prowadził badania terenowe na obszarze ukraińskich gwar Pobuża oraz gwar litewskich w powiecie sejneńskim. Jest współautorem dwóch tomów *Gramatyki konfrontatywnej bułgarsko-polskiej* oraz syntezy *Polsko-bułgarskiej gramatyki kontrastywnej*, autorem dwóch monografii poświęconych semantycznej kategorii określoności/nieokreśloności oraz modalności imperceptywnej w językach polskim i litewskim (badania kontrastywne). Jest/był kierownikiem i/lub wykonawcą kilkunastu projektów badawczych/grantów (w tym międzynarodowych). TW wolnych chwilach tłumaczy prace naukowe, dokumenty prawne, prawnicze, specjalistyczną dokumentację techniczną (języki litewski, rosyjski, ukraiński, bułgarski). Jest redaktorem naczelnym „*Cognitive Studies | Études cognitives*”.

Kontakt: roman.roszko@ispan.waw.pl

NOTY O AUTORACH

TOMASZ SZUTKOWSKI

Dr hab., prof. US, językoznawca, pracownik Instytutu Językoznawstwa Uniwersytetu Szczecińskiego, kierownik studiów podyplomowych Szkoła Tłumaczy i Specjalistycznych Języków Obcych, członek Zespołu Badawczego Frazeologii i Paremiologii, autor trzech monografii naukowych, współredaktor kilku opracowań zbiorowych i autor ponad pięćdziesięciu prac z zakresu onomastyki, frazeologii, paremiologii, leksykografii i translatoryki.

Kontakt: tszutkowski1@wp.pl

ALEKSANDRA SZYMAŃSKA

Doktor nauk humanistycznych, adiunkt w Zakładzie Literatury i Kultury Rosyjskiej przy Instytucie Rusycystyki UŁ. Autorka monografii *Postać Don Juana w utworach pisarzy rosyjskich XIX wieku* (2009) oraz artykułów dotyczących związków literatury rosyjskiej z hiszpańską w zakresie dziedziczenia mitów, jak również ich ewolucji i specyfiki rosyjskiej. Zainteresowania naukowe: historia literatury rosyjskiej XIX–XXI wieku, intertekstualność jako problem teoretyczno-metodologiczny we współczesnym literaturoznawstwie rosyjskim, semiotyczna koncepcja kultury, motywy hiszpańskie w literaturze rosyjskiej.

Kontakt: aleksandra.szymanska@uni.lodz.pl

JELENA WAŻENINA

Dr nauk filologicznych, adiunkt w Katedrze Języka Ukraińskiego, Teorii i Historii Literatury Ukraińskiej i Światowej Donieckiego Uniwersytetu Narodowego imienia Wasyla Stusa (Winnica). Zainteresowania naukowe: język ukraiński, stylistyka, frazemostylistyka, frazemodydaktyka, etnolingwistyka, słowotwórstwo języka ukraińskiego, metodyka nauczania języka ukraińskiego.

Kontakt: o.vazhenina@donnu.edu.ua

ANATOLIJ ZAHNITKO

Dr hab. nauk filologicznych, profesor na Wydziale Lingwistyki Ogólnej i Stosowanej oraz Filologii Słowiańskiej Donieckiego Uniwersytetu Narodowego imienia Wasyla Stusa (Winnica). główny pracownik naukowy Ukraińskiej Fundacji Językowo-Informacyjnej NAN Ukrainy, członek korespondent NAN Ukrainy. Zainteresowania naukowe: językoznawstwo teoretyczne, ogólne, porównawczo-historyczne i słowiańskie, funkcjonalne, komunikacyjne, tekstowe, korpusowe, stosowane; teoria systemów językowych; gramatyka przypadkowa, analityczna, formalna. Monografie: *Teorija linhwopersonoohiji* (2017); *Mownyj prostir hramatyky* (2018); *Delineation of Linguopersonology and Linguoaxiology* (we współautorstwie, 2019); *Teoriji linhwistycznych uczeń* (2019); *Współczesny słownik lingwistyczny* (2020).

Kontakt: a.zagnitko@donnu.edu.ua

ALEKSANDRA ZIĘBA

Mgr, absolwentka Wydziału Filologicznego Uniwersytetu Śląskiego w Katowicach, kierunek: język rosyjski, specjalność: tłumaczeniowa. Obecnie doktorantka w Szkole Doktorskiej Uniwersytetu Śląskiego w dyscyplinie: językoznawstwo. Zainteresowania badawcze: leksykografia przekładowa, lingwistyka korpusowa, słotwórstwo, aspekt motywacyjny, język rosyjski, język czeski, język hebrajski. Autorka kilku prac z zakresu językoznawstwa rosyjskiego, polko-rosyjskiego oraz przekładoznawstwa.

Kontakt: aleksandrazieba33@gmail.com