



MARIA MOCARZ-KLEINDIENST

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

Katolicki Uniwersytet Lubelski Jana Pawła II, Lublin

KORPUSY JĘZYKOWE W POLSKO-ROSYJSKIM 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. Kilgariff, V. Baisa, J. Bušta, M. Jakubiček, 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. Faaß, *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 geograficzne 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 odpowiedniości 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 2013; 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)

Identyfikator	Fragment tekstu	Wyraz bazowy	Terminy powiązane
1.	Przenikniony zbież zawięz 240,192,461 słów. Znalaziono 382 akapity najmniej do zapytania w 0,063s. Bieżąca strona zawiera 101 przykładów z 58 różnych tekstów.	elewacja	Wanna, Muzany, Suwa...
2.	plechonia z XIV w. Kościół (dawnej szpitalny) z 1 poł. XVII w.	elewacja	frontona i skłapy z 1723.
3.	8x1,5 m, co daje szerokości wiązki ok. 0,90 (argument) i 4,50	elewacja), Każdy z falowców stanowi wierz elementów promieniujących i
4.	amnaty i warunki meteorologiczne, natężenia kątowno amnaty (argument)	elewacja) w taki sposób, by zapewnić możliwość dlate prawopodobnie
5.	pręta środkowa z głównym wejściem. Od strony ogrodu natomiast	elewacja	palacu została oprowadana nadle oryginalnie: potrodoż zabył ja
6.	jest to piec kachelowy, dwukondygnacyjny, zwieszony - podobnie jak	elewacja	palacu - krenelałem, z pamiłkami i gotyckimi sterczykami. To
7.	Drugą akwarię ukazuje nam palec od strony ogrodu. Tu	elewacja	jest plaaka, pozbawiona ryzalitów, jedynie poordło na parterze
8.		Elewacja	ogrodowa natomiast, akcentowana potrodoż północnym ryzalitem.
9.		Elewacja	palacowa, zaryta przez powien cna białym nie tamięję już
10.	gdy podchodzimy bliżej, okazuje się, że cna węgłna dwukondygnowa	elewacja	nie jest główną fasadą palacu, a jedynie jego boczną, zachodnią
11.	się do palacu, stopniowo ukazuje się naszym oczom frontona	elewacja	- Najtjierw dostrzegamy dziewiczościową, symetryczną sylwetkę
12.	do uzyskania jednolitości w ramach jednego układu przestrzennego	elewacja	frontona), a z drugiej, przeciwstawienie sobie odmiennych i
13.	sobie odmiennych i poziomie do siebie nie panujących ścianek (elewacja	- galerie).
14.	sie o tym opowiadają już pod dachem palacu. Najtjierw jednak ziga	elewacja	ogrodowa.
15.	w swym wnętrzu, jak to zarysują ławy, salony. Pora tym ryzalitem	elewacja	ogrodowa przedstawia się raczej dość skromnie, a nawet bardzo
16.	przedstawia się raczej dość skromnie, a nawet bardzo skromnie.	Elewacja	nie posiada żadnych elementów wertykalnych, a działa układem
17.	z herbami "Nalecz" i "Drogosław". A zatem to jest właściwa	elewacja	frontona palacu w Dobrzycu.
18.	obliczonej na grę wiwiala i cienia, fasady palacu, jego zachodnią	elewacja), podobnie jak i ta od strony wschodu, prezentuje się bardzo
19.	latając wiwiał jak na dachu aż do końca Prusa. Dama uprwa	elewacja	kamienne. A kosciołek to już zwyczajna...
20.	dwoma łupkami i prostokątną zakrytą. Interesująca barokowa	elewacja	oszczędz wieża z niskim hełmem i latarnią.
21.	z późniejszych okresów, m. in. na narożnej pod nr. 3 - barokowa	elewacja), inne klasycystyczne lub eklektyczne. We wach, pierzei Rynku na
22.	świątkowa barokowo-klasycystyczna wykonana w l. 1785-80) oraz	elewacja	czelowa wraz z wieżą o kopułowym hełmie. We wnętrzu sklepiana
23.	w Nowej prechleńskim kryje sięklepianym katedrowym z łaciennymi.	Elewacja	wschodnią (wsparty strypami barokowa dekoracją podobnie jak
24.	wieża ewangeliczna barokowymi łaciennymi z latarniami.	elewacja	podzielona jest wydatnym gzymsem, wólcob którego grobiąca rząd
25.	jenitów było kolegium (1669-86), na planie litery E, którego	elewacja	czelowa wzorowana jest na palacach i kolegiach w Pradze i Hradcu
26.	w. Na piętrze gotycka sala zebrań o sklepieniu siatkowym. Ciekawa	elewacja	o gotyckim szczycie, woparta sukkowymi szkarpmi.
27.	W pięknym parku partnersowy palec z XVIII w., przebudowany w XX w.	Elewacja	złożona barokowym ornamentem i palastami. We wnętrzu zażykowy
28.	z ul. Brzozowa, a lewej, fragmenty gotyckie, m. in. cokolion	elewacja	"domu gotyckiego" z XIV w., zapewnie spikieralny. Jedną z
29.	Powstał on w XVI-XVII w. Widoczna z niego ceglana gotycka	elewacja	średnio-wiecznego Domu Wielkiego (XV w.), zdołona ślepną
30.	wejście do parku królewskich. Najbardziej jest późnobarokowa	elewacja	skrzydła pa. z Brana Senatorską (1742, A. Solari). Fasada zach. z
31.	Fasada zach. z Wieżą Zygmontowską utrzymaną w stylu XVII w.	Elewacja	skrzydła pd. (1767, J. Fontana) późnobarokowa.
32.		Elewacja	osobliwa dekoracja typu signifio. W otoczeniu park z poz.

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 użyc 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 użyc w KZ, 1 – w KO), *zachodnia elewacja* (2 użyc w KZ, 8 – w KO), *elewacja zachodnia* (1 użycie w KZ, 27 – w KO) *barokowa elewacja* (2 użyc 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 haseł. Potwierdza to następujący przykład:

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

Dane tekstu dla wyniku:	4
Identyfikator tekstu:	IPPAN_p0000100137
Tytuł:	<i>Cenne zbory zwolich, wstaje nielice</i>
Wydawca:	Przebiehienstwo Wydawnicze Produkcynie Ushgowe Sp. z o.o.
Zeszyt:	Typodruk Regionalny "Gazeta Ciepochowska" nr 16 (80)
Kanał:	Hasła jęzika Tygodnik
Typ:	#typ_pabl
Autorzy:	UG
Data publikacji:	2007-04-19
<p>Dziękuję! Kategorie Sockakowicza. "Budynek neopotyka z olennych aobokowych, pocz. XX w. (przed 1914). Dwupiętrowy z mieszkalnymi interesami, pięta dziesięciokątne, parter trzynastokątny, z siania przelazdową na szóstej osi, kryta posreżym sklepianem okolicznym z zależnym rehowaniem o składowie krzyżowym, z gurtami spływającymi na przyciemne pilastry. Wnętrze czepkowo przekształcone, z zakładowymi meblami osmaleni. W klatce schodowej drzwi do mieszkalni w podziemiach pływających, z osobnymi napowietrznymi. Balkonami schodów zochonkowa, zildana, o motowych egwos-skalowych, od frontu analogiczne balkonady, przy schodach do latarni. Elewacja frontowa z posreżymi ryzalitami skrajnymi oraz dwukondygnacyjnym wykuszem na osi środkowej, ponad korynnymi trójkątne uszczyty z mansardami - boniowana, z dekoracją neopotyka w formie laskowan (objmujących między innymi okna), sterczy i Fryzów arkadowych, zwielczona kamionami murekowaną balkonadą z ryzalitami balkonowymi w ryzalitech skrajnych, nad oknami drugiego piętra tace w otoczeniu neopotyka skrajnych. Brama wejściowa o kantolonych węgach z zależnymi ochronnymi ryzalitami i półkolistymi odbojami, w niej drzwi żelazne dwukrzydłowe z zakładowymi przyciemkami o motowych geometrycznych i murekowanych z ryzalitami. Czepkowo zachowana pierwotna stolarka. Dach mansardowy z wystawkami. "Książki zabytków smaku. Miasto Ciepochowa", pod red. Zofii Roznowy i Ewy Smalukowicz, cz. 1, Instytut Sztuki Polskiej Akademii Nauk, Warszawa 1995".</p>	

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-

zeniu 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. Hebal-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 cytataми 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 stalлы тонули в туманной полутьме.
— W kościele sam luksus — Drosselbart wyłamał z <i>empory</i> w prezbiterium kolejną deskę.	— В церкви сплошная роскошь. — Дроссельбарт выломал из галереи в пресвитерне очередную доску.

Powyższe 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 perspektywę przesiedlenia 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 stalle tonęły w mętym półmroku [<i>Narrenturm</i> , Sapkowski Andrzej] (NKJR).	Остальная часть пресвитерии , весь неф, деревянные эмпоры и сталлы тонули в туманной полутьме.
W stronie prezbiterium utworzyła się zbita masa «przystępujących» do antalka z winem [<i>Popioły</i> , Żeromski Stefan] (KPRUW).	У алтаря образовалась плотная толпа солдат, «со страхом и верою приступавших» к бочонку с вином [<i>Пепел</i> , Жеромский Стефан].
Nie pozostało nic innego, tylko wrócić do prezbiterium i szukać tam pożywienia. [<i>Popioły</i> , Żeromski Stefan] (KPRUW).	Цедро не оставалось ничего другого, как вернуться в алтарь и поискать там какой-нибудь снеди [<i>Пепел</i> , Жеромский Стефан].
Weszli do ciemnej kruchty i Witek zobaczył przy końcu nawy, pod balaskami prezbiterium , dwa rzędy płonących świec [<i>Kronika wypadków miłosnych</i> , Konwicky Tadeusz] (KPRUW).	Вошли в темный притвор, и Витек увидал в глубине нефа, под сиянием пресбитерия , два ряда горящих свечек [<i>Хроника любовных происшествий</i> , 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 wariantywność 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-

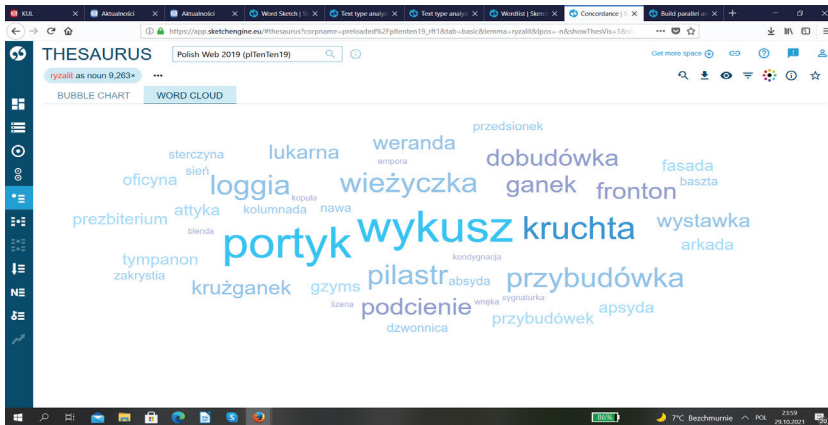
¹⁷ www.sketchengine.eu

¹⁸ Szerzej na temat platformy w pracy A. Kilgarriff, *The Sketch...*, s. 7–36, także: E.В. 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

tej oferty gotowych korpusów niekwestionowaną zaletą tej platformy jest dostępność narzędzi do budowania, wgrывania i wykorzystywania m.in. właśnie na potrzeby leksykografii własnych korpusów. Do pracy z językiem polskim Sketch Engine oferuje gotowy subkorpus Polish Web 2019 (plTenTen19) oraz OPUS 2 Polish. W tym drugim dominują materiały filmowe z platformy Open Subtitles (ponad 92%). Dla przeszukiwań w języku rosyjskim wykorzystano Russian Web 2011 (rusTenTen12) oraz OPUS 2 Russian z kolekcją tekstów MultiUN. Tworzy on zbiór przetłumaczonych dokumentów Organizacji Narodów Zjednoczonych, stanowiących 87% wszystkich tekstów. Zdecydowanie mniejszą kolekcję stanowią tu napisy filmowe OpenSubtitles2011 (8,75%). Sketch Engine umożliwia automatycznie generowane listy synonimów lub słów należących do tego samego pola semantycznego dzięki opcji Tezaurusu. Po wybraniu jej w lewym menu (wyróżnionym kolorem niebieskim) i wpisaniu słowa kluczowego można uzyskać zestaw leksemów (w danym wypadku ze zbioru polTenTen19) będących w bardzo bliskich lub luźnych powiązaniach semantycznych z leksemem wyjściowym. Ilustruje to poniższy przykład z użyciem terminu *ryzalit*:

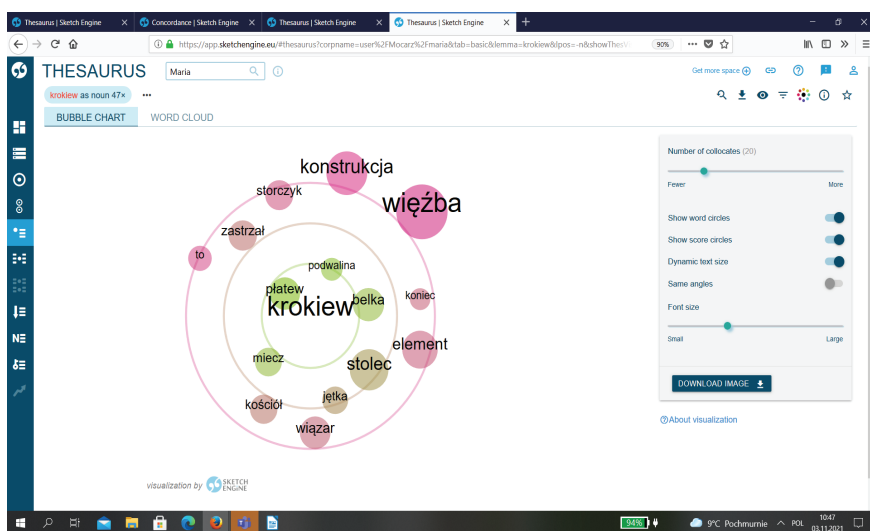
Tabela 7. Przykład tezaurusu dla terminu *ryzalit* (źródło: www.sketchengine.eu)



Po wpisaniu w rubryce w lewym górnym rogu wyszukiwanej frazy w tzw. chmurze wyrazów pojawia się zbiór jednostek semantycznie powiązanych z daną frazą. Zatem tezaurus w takiej postaci może być bardzo pomocny w tworzeniu bazy haseł wyjściowych. W danym

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 tezaury 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 tezaury znalazły zastosowanie w procesie konstruowania siatki hasłowej. Na ich podstawie do korpusu haseł włączono następujące bazowe terminy (oznaczające elementy konstrukcji dachowych): *zastrzał* (33 użyć w subkorpusie), *stolec* (76), *jętka* (25), *wiązar* (41), *miecz* (28), *storczyk* (32), *wieżba* (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 *wieżbza*:

Tabela 9. Możliwości dystrybucyjne terminu *wieżbza* na bazie własnego subkorpusu (źródło: www.sketchengine.eu)

a_modifier	prec_prep	post_verb	prec_verb	post_o	is_obj4
jętkowy Wieżbzy jętkowe	w w wieźbach	spotykać wieźb spotykamy	nazywać nazywamy wieźbą	stolec wieźby o stołach	wykształcić wieźby owarze wykształczy
otwarty otwiera wieźbę	przy Przy wieźbach	mieć wieźbach żyłowych mamy	tworzyć tworzy wieźby płatwiowo-kleszczowe	redukcja wieźb o redukcji	opuszciliśmy wieźby krótkowej opuszczimy
storczykowy wieźba storczykową zredakowaną	od od wieźb	nazywać wieźbe będziemy nazywali	wykonstruować wykonstruować wieźbę	przekrój wieźby o mniejszych przekrojach	pozwałać wieźby wolnej pozwala
taki W takich wieźbach	obok Obok powyższych wieźb	upodobniać wieźby upodobniają	pokazywać pokazuje wieźbę		nazywać wieźbę będziemy nazywali
płatwiowo-kleszczowych pokrępa wieźby płatwiowo-kleszczowych	jak jak wieźba	wykształcić wieźby krótkowej wykształczy	składać składa się wieźba		
płatwiowo-kleszczowe Wieźby płatwiowo-kleszczowe	do do wolnych wieźb	opuszciliśmy wieźby krótkowej opuszczimy	istnieć istnieć także wieźba		
dachowy Wieźby dachowe	nad nad wieźbami	polegać wieźbami wieszarowych polega	nazwać nazwać wieźbę		
półotwarty wieźby półotwarte	o o wieźbach	okazać wieźb okazały			
płatwiowo-kleszczowe Wieźbę płatwiowo-kleszczową o stołach		stanowiąc wieźby płatwiowo-kleszczowych stanowią			
wolny Wieźby wolne		pozwałać wieźby wolnej pozwala			
wieszarowy wieźbach wieszarowych		składać Wieżba składa			
płatwiowo-kleszczowe wieźby płatwiowo-kleszczowe		tworzyć wieźby płatwiowo-kleszczowych tworzy			
jednostorczykowa wieźba jednostorczykowa					

Narzędzie to pozwoliło na wyszukanie i umieszczenie w strukturze siatki hasłowej jako podhasła następujących terminów złożonych o strukturze imiennej: *wieżbza dachowa jętkowa*, *wieżbza storczykowa*, *wieżbza stolcowa*, *wieżbza płatwiowo-kleszczowa*, *wieżbza otwarta*, *wieżbza wieszarowa*, *wieżbza 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. *wieżbza jętkowa*. Niemniej podejmowane próby poszukiwań ekwiwalentów mają wartość sondazową, podobną do tej, jaka została zaprezentowana wcześniej przy analizie przydatności zasobów NKJR. Ponownie wykorzystany autorski zbiór kolokacji z komponentem bazowym *φacad* 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-Jezierska 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 Jakubíček, 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)