



PIOTR ŻMIGRODZKI

ORCID: 0000-0001-6037-5507

Instytut Języka Polskiego Polskiej Akademii Nauk w Krakowie

piotr.zmigrodzki@ijp.pan.pl

## *Wielki słownik języka polskiego PAN* jako słownik tematyczny

### Słowa kluczowe

leksykografia polska, słowniki tematyczne, *Wielki słownik języka polskiego PAN*

### Key words

Polish lexicography, thematic dictionaries, *Wielki słownik języka polskiego PAN*

Jedną z cech słownika drukowanego, w przeciwieństwie do słownika elektronicznego, jest to, że informacje w nim zawarte można odczytywać w zasadzie tylko w jeden sposób, taki jaki zaplanowali jego autorzy i jaki wynika z jego zapisanej na papierze makro- i mikro-struktury. Przeważającą grupę słowników papierowych stanowią te w układzie alfabetycznym, a bywają również na przykład w układzie pojęciowym lub gniazdowym. Aby stworzyć czytelnikowi dodatkowe możliwości wyszukiwania informacji, w słownikach papierowych bywają dodawane indeksy, na przykład indeks rzeczowy do słownika w układzie alfabetycznym, indeks alfabetyczny do słowników w innym porządku (np. słownika synonimów). Możliwości korzystania ze słownika elektronicznego są natomiast znacznie większe i bardziej różnorodne, choć także zależą od sposobu opracowania materiału i jego udostępnienia. Jeśli tylko twórcy słownika oraz inżynierowie projektujący jego strukturę informatyczną to przewidzieli, hasła można przeglądać i wyszukiwać według rozmaitych kryteriów, wyświetlać tylko hasła spełniające określone warunki itd. To sprawia, że traci znaczenie, doniosła w erze słowników papierowych, problematyka typologii słowników, gdyż ten sam materiał leksykograficzny może być prezentowany w różnych kształtach i układach w obrębie jednego dzieła leksykograficznego.

Przykładem słownika polszczyzny opracowywanego od podstaw całkowicie w wersji elektronicznej, i tylko w takiej wersji udostępnianego, jest *Wielki słownik języka polskiego PAN* (dalej: WSJP PAN)<sup>1</sup>. Był on planowany jako słownik ogólny i jako taki jest zwykle rozpatrywany (stawia się go w szeregu takich dzieł, jak słowniki pod redakcją W. Doroszewskiego, M. Szymczaka czy S. Dubisza). Jednak fakt, że stanowi on przeszukiwalną bazę danych leksykograficznych, powoduje, iż można z niego korzystać w różny sposób, na przykład wyszukując i przeglądając tylko hasła frazeologiczne albo też sortować hasła według rozlicznych kryteriów, w tym według klasyfikacji tematycznej zawartego w nim słownictwa. Przybliżenie tej funkcji WSJP PAN jest pierwszorzędnym celem niniejszego artykułu. Cel dodatkowy to informacja o statystycznej strukturze pól tematycznych leksyki opisanej w słowniku i przedstawienie wstępnych wniosków, jakie z tych zestawień liczbowych płyną.

Potrzeba publikacji tak zaplanowanego tekstu wynika z przeświadczenia, że przydatność WSJP PAN jako słownika tematycznego jest wśród odbiorców (których liczba, szacowana na podstawie odwiedzin witryny internetowej, wynosi przynajmniej 200 tys. tzw. unikalnych użytkowników miesięcznie) słabo uświadamiana albo też w ogóle nieuświadamiana. Za koronny przykład może tu posłużyć znamieny fakt. Najwybitniejsza polska znawczyni zagadnień badawczych słownictwa tematycznego, Zofia Cygal-Krupa, w przeglądowym artykule z 2017 roku, w którym sporo miejsca poświęciła omówieniu tematycznego podziału słownictwa zastosowanego w WSJP, zawarła następującą konkluzję: „chciałabym sformułować postulat opracowania w pełni nowoczesnego słownika tematycznego języka polskiego jako pierwszorzędnego zadania stojącego przed polskim językoznawstwem”<sup>2</sup>. Jasne jest więc, że badaczka nie postrzega WSJP PAN jako słownika tematycznego. Wynika to zapewne z jej przywiązania do obrazu sztywnej makrostruktury słownika i nieświadomości, że WSJP PAN – kiedy użyjemy odpowiedniej funkcji wyszukiwania zaawansowanego – staje się słownikiem tematycznym, w dodatku większym i lepszym niż inne, bo zawierającym nie tylko same wykazy wyrażen przypisanych do danego pola tematycznego, ale też ich pełne opisy.

Tak postrzegany WSJP PAN jest największym słownikiem tematycznym współczesnej polszczyzny, przewyższającym o rząd wielkości inne dostępne. Według stanu na 15 sierpnia

<sup>1</sup> *Wielki słownik języka polskiego PAN* jest publikowany w internecie sukcesywnie od 2009 r. i dostępny bezpłatnie pod adresem <http://wsjp.pl>. Obejmuje on swym zasięgiem źródłowym polszczyznę od 1945 r. W sierpniu 2021 r. liczba opracowanych haseł przekroczyła 83,5 tys., a łączna objętość tekstu słownika w przeliczeniu na druk jest równoważna ok. 25 tomom objętości tomów słownika pod red. Witolda Doroszewskiego. Więcej informacji o słowniku można znaleźć na wskazanej stronie internetowej oraz w publikacji: *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*, red. Piotr Żmigrodzki i in. (Kraków: Instytut Języka Polskiego PAN, Wydawnictwo Libron, 2018).

<sup>2</sup> Zofia Cygal-Krupa, „Słownictwo tematyczne w teorii i praktyce językoznawstwa polskiego”, w: *Terminologia specjalistyczna w teorii i praktyce językoznawców słowiańskich*, red. Renata Przybylska, Władysław Śliwiński (Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2017), 116.

2021 roku<sup>3</sup> miał on 141 561 znaczeń sklasyfikowanych pod względem pól tematycznych. Liczba ta przekracza wskazaną wcześniej ogólną liczbę haseł w słowniku, gdyż klasyfikuje się odrębnie każde z wyróżnionych znaczeń w haśle (a są w słowniku artykuły hasłowe opisujące po kilka, nawet kilkanaście znaczeń). Klasyfikacji podlegają wszystkie hasła oprócz reprezentujących tzw. jednostki funkcyjne (a więc m.in. przyimki, spójniki, partykuły i inne). Te są z klasyfikacji tematycznej *ex definitione* wyłączone, gdyż nie odnoszą się do rzeczywistości pozajęzykowej (a to jest istotą klasyfikacji tematycznej), ale sygnalizują relacje pomiędzy składnikami tekstu. Aby uzmysłowić sobie przewagę liczbową WSJP PAN nad dostępnymi słownikami tematycznymi polszczyzny, można podać przykładowo, że słownik w książce Andrzeja Markowskiego *Leksyka wspólna różnym odmianom polszczyzny* (1992) liczy 6 tys. jednostek, *Ilustrowany słownik podstawowy języka polskiego* Zofii Kurzowej (1999) – 5 tys., *Słownik tematyczny języka polskiego* Małgorzaty Kity i Edwarda Polańskiego (2002) – najwyżej kilkanaście tysięcy (brak jest konkretnych danych), *Wielki słownik frazeologiczny* Renarda Lebdy (2009), alfabetyczny, ale z indeksem tematycznym – 18 tys. Znacznie bogatszy jest elektroniczny *Słownik pojęciowy języka staropolskiego* pod red. Bożeny Sieradzkiej-Baziur, który liczy około 46 tys. znaczeń sklasyfikowanych tematycznie<sup>4</sup>, nie dotyczy on jednak polszczyzny dzisiejszej. Podobną liczbę jednostek notują największe słowniki przekładowe w układzie tematycznym.

Klasyfikacja tematyczna w WSJP PAN jest wynikiem ogromu pracy, zarówno koncepcyjnej, jak i ściśle operacyjnej, leksykograficznej. Jej podstawy teoretyczne i szczegóły techniczne wraz z zasadami przydziału jednostek do wyodrębnionych grup tematycznych opracowała Barbara Batko-Tokarz. Zostały one wyczerpująco opisane wraz z uzasadnieniem metodologicznym w jej monografii<sup>5</sup>, do której należy odesłać zainteresowanego czytelnika<sup>6</sup>. Skrótowo mówiąc, klasyfikacja opiera się na antropocentrycznym, nie zaś scjentyistycznym oglądzie świata, co ją przybliży na przykład do klasyfikacji wcześniej opracowanej przez Andrzeja Markowskiego, a różni od starszych podziałów, przede wszystkim tego wprowadzonego przez Marka Rogeta w jego słynnym teaurusie, publikowanym od połowy XIX w.<sup>7</sup> czy Rudolfa Halliga i Walthera von Wartburga<sup>8</sup> (z której korzystają

<sup>3</sup> Wszystkie podawane w artykule liczby dotyczą stanu bazy słownika z tego właśnie dnia.

<sup>4</sup> Por. Bożena Sieradzka-Baziur, *Onomazjologiczny słownik on-line. Metodologia, zawartość, wykorzystanie w badaniach na przykładzie Słownika pojęciowego języka staropolskiego* (Kraków: Instytut Języka Polskiego PAN, 2020).

<sup>5</sup> Barbara Batko-Tokarz, *Tematyczny podział słownictwa współczesnego języka polskiego. Teoria, praktyka, leksykografia* (Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2019).

<sup>6</sup> Bardziej zwięzła wersja opisu znajduje się w rozdziale poświęconym klasyfikacji tematycznej, zamieszczonym w książce przywołanej w przypisie 1 niniejszego artykułu, a także, oczywiście, w zasadach opracowania WSJP PAN dostępnych na jego stronie internetowej.

<sup>7</sup> *Dr Roget's Thesaurus of English Words and Phrases classified and arranged to facilitate the Expression of Ideas and assist in Literary Composition* (London 1852).

<sup>8</sup> Rudolf Hallig, Walther von Wartburg, *Begriffssystem als Grundlage für die Lexikographie* (Berlin: Akademie-Verlag, 1952).

chętnie polscy diachroniści). Podział jest trójstopniowy: na pierwszym poziomie mamy siedem kategorii, nazwanych przez autorkę sferami; ich nazwy już dobitnie wskazują na punkt widzenia klasyfikacji:

- I. CZŁOWIEK JAKO ISTOTA FIZYCZNA.
- II. CZŁOWIEK JAKO ISTOTA PSYCHICZNA.
- III. CODZIENNE ŻYCIE CZŁOWIEKA.
- IV. CZŁOWIEK W SPOŁECZEŃSTWIE.
- V. CZŁOWIEK I TECHNIKA.
- VI. CZŁOWIEK I PRZYRODA.
- VII. KATEGORIE FIZYCZNE.

W obrębie każdej sfery wyróżniane są następnie pola (w ogólnej liczbie 45), a w ich obrębie podpola (w liczbie 205). Przykładowo, sfera I dzieli się na pola: 1. Budowa i funkcjonowanie ciała ludzkiego; 2. Bieg życia; 3. Pielęgnacja ciała; 4. Choroby i ich leczenie; 5. Określenia fizyczności człowieka, z kolei np. w obrębie pola 1. Budowa i funkcjonowanie ciała ludzkiego wyróżniamy podpola: 1.1. części ciała, elementy i substancje składowe; 1.2. czynności i stany fizjologiczne; 1.3. seksualność; 1.4. zmysły. Z pełnym układem kategorii można się zapoznać w przywołanej monografii B. Batko-Tokarz, a także w *Zasadach opracowania* słownika, dostępnych na jego stronie internetowej.

Samym przydziałem znaczeń do pól tematycznych od początku prac do 2018 roku zajmował się wydzielony zespół osób (w terminologii przyjętej w zespole WSJP i używanej w publikacjach – specjalistów) pod kierunkiem B. Batko-Tokarz. Później, w związku ze zmniejszeniem zespołu wymuszonym względami finansowymi, rolę tę przejęli sami redaktorzy artykułów hasłowych. Większość znaczeń w słowniku jest sklasyfikowana z dokładnością do trzeciego poziomu klasyfikacji, bywają jednak hasła, w których przyporządkowanie ogranicza się tylko do drugiego poziomu lub nawet tylko do pierwszego. Dotyczy to haseł o bardzo ogólnym znaczeniu, np. jednostka leksykalna *człowiek* 1 ‘istota ludzka’ przypisana została do dwóch pierwszych sfer podziału (CZŁOWIEK JAKO ISTOTA FIZYCZNA i CZŁOWIEK JAKO ISTOTA PSYCHICZNA), bez doprecyzowania klasyfikacji na kolejnych poziomach. Dane znaczenie może być włączone do więcej niż jednej grupy słownictwa tematycznego. Liczba takich wypadków jest stosunkowo pokaźna. Spośród wszystkich sklasyfikowanych 141 561 znaczeń 38 215 było umieszczonych jednocześnie w dwóch polach, a 4001 – w trzech i więcej. Jest też pewna, stosunkowo niewielka, liczba znaczeń uznanych za „atematyczne”, czyli takie, które nie wykazują cech pozwalających je zaliczyć do jednego lub kilku, nawet kilkunastu konkretnie wskazanych pól. Za typowy przykład można uznać czasownik *być* w jego funkcji słowa posiłkowego. Kategorie tematyczne są przypisywane hasłom oczywiście dopiero po ustaleniu ostatecznej postaci istotnych elementów ich mikrostruktury, takich jak przede wszystkim definicja semantyczna, kwalifikatory, kolokacje, cytaty. Ostateczne zaklasyfikowanie tematyczne jednostki leksykalnej jest rezultatem łącznego rozpatrzenia tych wszystkich elementów.

Na koniec tej części rozważań warto jeszcze zwrócić uwagę na różnice między klasyfikacją tematyczną słownictwa w słowniku a zastosowaniem kwalifikatorów specjalistycznych, takich jak *med.*, *fiz.*, *techn.* itd. Te ostatnie wskazują na przynależność jednostki leksykalnej do specjalistycznej odmiany językowej (profesjolektu, ewentualnie socjolektu), kategorii klasyfikacji tematycznej zaś – do zasobu słownictwa o konkretnej tematyce. W wyniku zastosowania jednej i drugiej klasyfikacji otrzymujemy zbiory nakładające się. Na przykład w obrębie pola 4.1. Rodzaje chorób znajdujemy jednostkę *alzheimer*, zakwalifikowaną jako przynależną do odmiany potocznej, jej zaś synonim, *choroba Alzheimera* odnotowany jest z kwalifikatorem specjalistycznym *med.* Podobnie w obrębie pary: *dalekowzroczność – nadwzroczność* tylko drugi wyraz ma kwalifikator *med.*, a pierwszy występuje bez kwalifikatora. Bez kwalifikatora *med.* odnotowane są przykładowo takie leksemy z podpola „rodzaje chorób”, jak *angina*, *grypa*, *gościec*, gdyż są one powszechnie używane w komunikacji niespecialistycznej. Jest to konstatacja ważna dla prawidłowego zrozumienia opisu leksykograficznego w WSJP PAN, ponieważ wiadomo, że w niektórych słownikach, w których klasyfikacji tematycznej nie ma, np. w PWN-owskim *Uniwersalnym słowniku języka polskiego*, kwalifikatory specjalistyczne intencjonalnie pełnią funkcję *quasi*-określników tematycznych, co sprawia, że opatruje się nimi także wyrazy nienależące do warstwy słownictwa profesjonalnego.

W aktualnej postaci WSJP PAN klasyfikacja tematyczna funkcjonuje w dwojaki sposób. Po pierwsze, jest ona widoczna przy każdym znaczeniu, które według niej sklasyfikowano, po otwarciu odpowiedniej zakładki widoku hasła lub też w widoku zbiorczym, co obrazują ilustracje:

The screenshot displays the online dictionary interface for the word "angina". At the top, there is a navigation bar with the logo "WS JP" and the text "Wielki słownik języka polskiego". To the right of the logo are links for "Historia projektu", "Autorzy", "Zasady opracowania", and "Bibliograf".

The main content area shows the word "angina" in a large, bold font. Below it is a hierarchical diagram illustrating the classification of the word:

```

    CZŁOWIEK JAKO ISTOTA FIZYCZNA
    ↓
    Choroby i ich leczenie
    ↓
    rodzaje chorób
  
```

At the bottom of this diagram are two buttons: "CHRONOLOGIZACJA" and "POKAŻ WSZYSTKO".

On the right side of the page, there is a vertical list of view options, each in a separate box:

- Definicja
- Kwalifikacja tematyczna
- Relacje znaczeniowe
- Połączenia
- Cytaty
- Odmiana
- Pochodzenie

At the bottom of the page, there is a footer with the text: "Data ostatniej modyfikacji: 26.08.2014".

## angina

Chronologizacja

1 połowa XIX w., Bańk

Odmiana

część mowy: *rzeczownik*  
rodzaj gramatyczny: *ż*  
*liczba pojedyncza liczba mnoga*

M. <i>angina</i>	<i>anginy</i>
D. <i>anginy</i>	<i>angin</i>
C. <i>anginie</i>	<i>anginom</i>
B. <i>anginę</i>	<i>anginy</i>
N. <i>anginą</i>	<i>anginami</i>
Ms. <i>anginie</i>	<i>anginach</i>
W. <i>angino</i>	<i>anginy</i>

Pochodzenie

*łac.* angina 'zapalenie gardła'

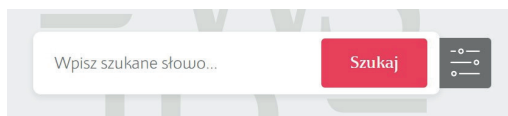
Definicja

choroba zakaźna objawiająca się stanem zapalnym gardła, powodującym ból i trudności w przełykaniu, oraz białym nalotem w gardle

Kwalifikacja tematyczna

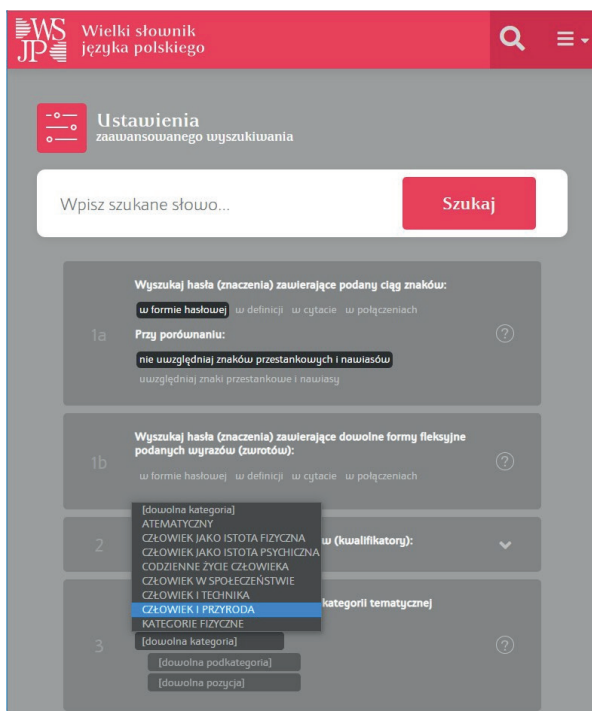
CZŁOWIEK JAKO ISTOTA FIZYCZNA → Choroby i ich leczenie → rodzaje chorób

Czytając hasło, można się więc zorientować, do jakich klas tematycznych jego znaczenie zostało przydzielone, nie jest to jednak w żadnym wypadku główna funkcja całej pracochłonnej operacji klasyfikowania tematycznego słownictwa<sup>9</sup>. Podstawowym jej celem jest umożliwienie wyszukiwania hasel (znaczeń) w słowniku według pól tematycznych. Obecnie można to uczynić, posługując się funkcją wyszukiwania zaawansowanego w WSJP PAN. Okienko wyszukiwania zaawansowanego można otworzyć, klikając szary przycisk na prawo od okienka wyszukiwarki głównej na stronie początkowej:



Otwiera się wówczas okno wyszukiwania zaawansowanego, w którym można wybrać różne kryteria przeszukiwania bazy słownika. Pod numerem 3 znajduje się drzewko wyszukiwania tematycznego; wybiera się najpierw poziom sfery (najwyższy, por. rysunek), a wybór ten pozwala odsłonić kolejne opcje menu i wejść w klasyfikacje coraz bardziej szczegółowe. Kryteria wyszukiwania można łączyć, np. szukając hasel z grupy „rodzaje chorób” będących rzeczownikami z kwalifikatorem *med.* i pochodzących z określonego języka.

<sup>9</sup> Najwyraźniej z braku zrozumienia tej kwestii wynikały w przeszłości krytyczne uwagi kierowane pod adresem twórców WSJP PAN. Przykładowo, podczas konferencji „Głosa do leksykografii polskiej” w 2011 r. wybitna i wpływowa warszawska lingwistka dobitnie wyraziła w dyskusji pogląd, że klasyfikowanie tematyczne słownictwa jest „w ogóle śmieszne i niepotrzebne”.



W odpowiedzi na kwerendę wyszukiwarka zwraca listę wyników, która ma postać form hasłowych z odsyłaczami do konkretnych haseł. Przykładowo w wypadku wyszukiwania: CZŁOWIEK JAKO ISTOTA FIZYCZNA – Choroby i ich leczenie – rodzaje chorób, początek listy wyników (obejmującej ogółem 750 pozycji) wygląda następująco:

- acydemia
- adidas
- aeroembolizm
- afatyczny
- afatyk
- afazja
- agorafobia
- AIDS
- aids
- aidsowiec
- aidsowy
- akrofobia
- alergia

Po kliknięciu danej formy hasłowej czytelnik zostaje przeniesiony do widoku ogólnego hasła. Jeśli hasło to ma więcej znaczeń, identyfikator znaczenia spełniającego warunki wyszukiwania wyświetla się na czerwonym tle, por. np. hasło *alergia*:

The screenshot shows the website interface for the 'Wielki słownik języka polskiego' (WSJP). At the top, there is a red header with the logo 'WS JP' and the text 'Wielki słownik języka polskiego'. Below the header, there are navigation links: 'Historia projektu', 'Autorzy', 'Zasady opracowania', and 'Bibliografia'. The main heading is 'alergia'. Below the heading, there is a sub-heading: 'Hasło ma wiele znaczeń, wybierz to, które Cię interesuje'. There are three search results listed: '1.a choroba', '1b wysypka', and '2. niechęć'. At the bottom of the search results, there are two buttons: 'CHRONOLOGIZACJA' and 'POKAŻ WSZYSTKO'. At the very bottom, there is a small text: 'Data ostatniej modyfikacji: 09.10.2013'.

Ostatnią kwestią do omówienia w niniejszym tekście jest struktura tematyczna słownictwa uwzględnionego w WSJP PAN. W tabeli 1 uszeregowano kategorie pierwszego poziomu klasyfikacji (siedem „sfer” oraz grupę znaczeń atematycznych) według liczby znaczeń do nich przydzielonych. Podane wartości sumują się do liczby 188 037; jest to ogólna liczba przypisań kategorii tematycznych w bazie słownika; większa niż liczba sklasyfikowanych haseł, gdyż, jak wspomniano, część znaczeń jest przydzielona do więcej niż jednej grupy.

Tabela 1. Kategorie pierwszego poziomu klasyfikacji

Sfera	Liczba znaczeń	Procent całości
II. CZŁOWIEK JAKO ISTOTA PSYCHICZNA	44 361	23,60
IV. CZŁOWIEK W SPOŁECZEŃSTWIE	41 712	22,18
VII. KATEGORIE FIZYCZNE	26 847	14,27
III. CODZIENNE ŻYCIE CZŁOWIEKA	25 309	13,46
I. CZŁOWIEK JAKO ISTOTA FIZYCZNA	18 600	9,89
VI. CZŁOWIEK I PRZYRODA	15 750	8,38
V. CZŁOWIEK I TECHNIKA	13 275	7,06
ATEMATYCZNY	2 83	1,16

Źródło: opracowanie własne.

Tabela 2 przedstawia odpowiedni ranking bogactwa słownictwa zebranego w „polach” (drugi stopień klasyfikacji). Liczby rzymskie odnoszą się do danej sfery (por. wyżej), numery arabskie zaś zgodne są z numeracją pól stosowaną w opisach klasyfikacji publikacji książkowej przywołanej wcześniej w przypisie 1.



Tabela 2. Ranking bogactwa słownictwa

Numer	Pole klasyfikacji tematycznej	Liczba znaczeń
1	2	3
VII. 42	Cechy i właściwości materii	19 759
II. 8	Ocena i wartościowanie	15 052
III. 23	Zasady współżycia społecznego	7 804
II. 9	Działalność intelektualna człowieka	7 418
II. 11	Relacje międzyludzkie	7 145
VI. 41	Świat zwierząt	6 998
III. 18	Praca	6 430
I. 1	Budowa i funkcjonowanie ciała ludzkiego	6 233
II. 10	Działalność artystyczna człowieka	6 081
III. 15	Jedzenie i jego przygotowanie	5 866
I. 4	Choroby i ich leczenie	5 270
IV. 19	Funkcjonowanie państwa	4 762
V. 33	Urządzenia i maszyny	4 278
VI. 40	Świat roślin	4 075
IV. 22	Prawo i łamanie prawa	3 650
IV. 25	Tradycja	3 561
I. 5	Określenia fizyczności człowieka	3 460
II. 2	Stany psychiczne człowieka	3 413
II. 7	Usposobienie człowieka	3 413
II. 24	Język	3 283
V. 36	Transport	3 279
VI. 39	Ziemia	3 263
V. 20	Przynależność i podział terytorialny	3 262
IV. 26	Religia – kościół	3 252
III. 13	Najbliższe środowisko życia człowieka	3 196
IV. 30	Finanse	3 184
III. 13	Ubranie	3 133
III. 16	Czas wolny	3 023
III. 17	Sport	2 655
IV. 21	Wojsko i wojna	2 640
VII. 43	Cechy i właściwości przestrzeni	2 619
VII. 44	Cechy i właściwości czasu	2 378
IV. 27	Edukacja i oświata	2 372
IV. 29	Biznes	2 354
I. 2	Bieg życia	2 203
V. 31	Budownictwo	2 170
V. 34	Przemysł	2 090
VII. 45	Ilość, liczby i liczenie	2 077
IV. 28	Środki masowego przekazu	1 506

1	2	3
3	Pielęgnacja ciała	1 382
II. 12	Rodzina	956
V. 32	Komunikacja na odległość	867
VI. 38	Pogoda	848
V. 35	Energia	542
V. 37	Niebo i wszechświat	535

Źródło: opracowanie własne.

Jak widać, najliczniej reprezentowane podpole wcale nie należy do najliczniej reprezentowanej sfery, co pośrednio świadczy o dużym zróżnicowaniu liczbowym słownictwa w poszczególnych grupach klasyfikacji.

Tabela 3 przedstawia dwadzieścia najliczniej reprezentowanych podpól (pozycji z trzeciego, najbardziej szczegółowego poziomu) klasyfikacji.

Tabela 3. Dwadzieścia najliczniej reprezentowanych podpól klasyfikacji

Numer	Podpole	Liczba znaczeń
8.2	wola, postawy, nastawienie człowieka wobec świata i życia	8957
11.1	określenia relacji międzyludzkich	6027
42.12	jakość i intensywność	5411
8.3	słownictwo oceniające	5268
42.11	ruch i spoczynek	4036
23.2	zachowania i wyrażenia nieakceptowalne społecznie	3739
9.1	nazwy działań intelektualnych człowieka	3706
18.1	rodzaje pracy	3230
23.3	stosunki, grupy i organizacje społeczne	3155
42.9	stan skupienia, trwałość materii	3025
1.2	czynności i stany fizjologiczne	2595
6.1	emocje i uczucia	2560
7.1	cechy charakteru, temperamentu, zdolności i umiejętności	2332
6.2	zachowania emocjonalne	2326
15.1	rodzaje pokarmu/potrav i ich cechy	1961
10.3	muzyka	1940
36.1	transport lądowy	1885
41.1	budowa i funkcjonowanie organizmów zwierzęcych	1849
5.1	wygląd	1759
4.4	leki i sposoby leczenia, zabiegi, sprzęty medyczne	1700

Źródło: opracowanie własne.

Warte prezentacji jest też zestawienie podpól najslabiej reprezentowanych w słowniku (tabela 4).

Tabela 4. Podpola najslabiej reprezentowane

Numer	Podpole	Liczba znaczeń
3.3	osoby związane z pielęgnacją ciała	17
3.4	miejsca z związane z pielęgnacją ciała	32
40.4	rośliny wodne	37
38.1	klimat	54
30.5	ubezpieczenia	69
31.5	osoby związane z budową	80
37.1	badanie i podbój kosmosu	87
17.5	miejsca związane ze sportem	91
18.2	miejsca pracy	94
41.10	mikroorganizmy	102
30.6	papiery wartościowe i akcje/udziały	104
34.3	osoby związane z przemysłem	108
34.1	rodzaje przemysłu	111
40.7	rośliny egzotyczne	122
32.1	poczta	131
42.3	wzory	148
41.6	inne zwierzęta lądowe	150
29.1	reklama i marketing	153
43.3	odległość	158
43.6	ciężar	160
35.2	źródła energii	160

Źródło: opracowanie własne.

Niniejsza prezentacja danych dotyczących klasyfikacji tematycznej w WSJP PAN ma charakter tylko demonstracyjny i również wstępny charakter mają niżej przedstawione wnioski. Potwierdza się ściśle antropocentryczny charakter leksyki języka polskiego, a ponadto jej nastawienie na psychikę i uczucia ludzkie, w mniejszym stopniu niż na fizyczność, cielesność. Największa liczba zanotowanych jednostek leksykalnych (znaczeń) dotyczy bowiem sfery II, tzn. kontaktów człowieka ze światem, jego percepcji rzeczywistości, procesów psychicznych, intelektualnych oraz interakcji międzyludzkich i społecznych (sfera IV). Człowiek widziany przez pryzmat haseł WSJP PAN jawi się jako istota myśląca i przede wszystkim czująca, mająca aktywny stosunek do zastanej rzeczywistości, którego nie waha się wyrażać. Mniej liczny jest (co różni opisywany zasób od wcześniej publikowanych słowników tematycznych) zasób dotyczący tzw. życia codziennego (sfera III), a także kontaktów człowieka z przyrodą czy techniką. W obrębie sfery III dość liczną grupę

stanowi słownictwo nazywające „zachowania i wyrażenia nieakceptowane społecznie”, co prawdopodobnie może być uznane za smutny znak czasu.

Niektóre dane liczbowe mogą zaskakiwać, na przykład bardzo mały zasób słownictwa związanego z pielęgnacją ciała (dwa pierwsze miejsca na liście „negatywnej”), wydawałoby się bowiem, że współcześnie tematyka ta jest bogato reprezentowana w różnych tekstach<sup>10</sup>. Przyczyn takiego stanu może być kilka. Przede wszystkim fakt, że słownik jest wciąż niepełny; mimo opracowania ponad 80 tys. haseł nie obejmuje licznych warstw słownictwa, a ze względu na sposób jego opracowywania (poczynając od leksemów najczęstszych w korpusach polszczyzny początków XXI w.) i podstawy materiałowe można się spodziewać, że pewne odmiany językowe czy kręgi tematyczne nie zostały jeszcze w nim należyście uwzględnione. Może więc z analizą struktury tematycznej słownictwa w WSJP PAN należałoby jeszcze poczekać na jego dalsze powiększenie. Dodatkowym problemem jest to, że uzyskane dane trudno jest odnieść do ustaleń badaczy dokonanych na bazie słowników istniejących, a to przede wszystkim ze względu na niewspółmierność podstawy materiałowej (jak wspomniano, WSJP PAN jest zbiorem znacznie większym od wcześniejszych słowników tematycznych polszczyzny). Szczegółowe różnice kategorii w poszczególnych klasyfikacjach wydają się tutaj problemem mniejszym. Tak czy inaczej, wykonanie wskazanego zadania wymagać będzie kompleksowych studiów, których wyniki z pewnością rozrosną się do rozmiaru monografii, a nie tylko jednego czy kilku artykułów.

Można sobie jednak wyobrazić szczegółowe studia i analizy, korzystające z dostępnych każdemu użytkownikowi możliwości funkcji wyszukiwarki zaawansowanej WSJP PAN, na przykład:

- charakterystyka (m.in. gramatyczna, pragmatyczna) słownictwa w poszczególnych polach i podpolach,
- analiza „łączenia się” kilku kategorii w charakterystyce pojedynczej jednostki leksykalnej,
- badanie konkurencji kategorii tematycznej z przynależnością do odmian językowych (wyrażoną kwalifikatorami),
- analizy przynależności tematycznej znaczeń reprezentujących poszczególne typy haseł (np. idiomy, przysłowia).

Nie jest to jednak zadanie dla autorów słownika, ale raczej dla innych lingwistów<sup>11</sup>.

<sup>10</sup> Zwróciła na to uwagę w dyskusji po ustnej prezentacji tych danych prof. Ewa Kołodziejek.

<sup>11</sup> Latem 2022 roku, kiedy niniejszy artykuł był już w druku, udało się wdrożyć odrębny interfejs słownika tematycznego, dostępny ze strony początkowej. Pozwala on czytelnikowi bardziej świadomie i bardziej wydajnie korzystać z walorów klasyfikacji tematycznej słownictwa w WSJP PAN.

## Bibliografia

- Batko-Tokarz, Barbara. *Tematyczny podział słownictwa współczesnego języka polskiego. Teoria, praktyka, leksykografia*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2019.
- Cygal-Krupa, Zofia. „Słownictwo tematyczne w teorii i praktyce językoznawstwa polskiego”. W: *Terminologia specjalistyczna w teorii i praktyce językoznawców słowiańskich*, red. Renata Przybylska, Władysław Śliwiński, 97–118. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2017.
- Hallig, Rudolf, Walther von Wartburg. *Begriffssystem als Grundlage für die Lexikographie*. Berlin: Akademie-Verlag, 1952.
- Kita, Małgorzata, Edward Polański. *Słownik tematyczny języka polskiego*. Warszawa: Wydawnictwo Edukacyjne Graf-Punkt, 2002.
- Kurzowa, Zofia. *Ilustrowany słownik podstawowy języka polskiego: wraz z indeksem pojęciowym wyrazów i ich znaczeń*. Kraków: Universitas, 1999.
- Lebda, Renarda. *Wielki słownik frazeologiczny*. Kraków: Krakowskie Wydawnictwo Naukowe, 2009.
- Markowski, Andrzej. *Leksyka wspólna różnym odmianom polszczyzny*. T. I–II. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego, 1992.
- Roget, Mark. *Dr Roget's Thesaurus of English Words and Phrases classified and arranged to facilitate the Expression of Ideas and assist in Literary Composition*. London 1852.
- Sieradzka-Baziur, Bożena. *Onomazjologiczny słownik online. Metodologia, zawartość, wykorzystanie w badaniach na przykładzie „Słownika pojęciowego języka staropolskiego”*. Kraków: Instytut Języka Polskiego PAN, 2020.
- Żmigrodzki, Piotr, Mirosław Bańko, Barbara Batko-Tokarz, Jakub Bobrowski, Anna Czelakowska, Maciej Grochowski, Renata Przybylska, Jadwiga Waniakowa, Katarzyna Węgrzynek, red. *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*. Kraków: Instytut Języka Polskiego PAN, Wydawnictwo Libron, 2018.

## *Wielki słownik języka polskiego PAN* jako słownik tematyczny

### Streszczenie

Artykuł przynosi prezentację cech *Wielkiego słownika języka polskiego PAN*, które umożliwiają korzystanie z niego jako słownika tematycznego polszczyzny (największego pod względem liczby sklasyfikowanych jednostek). W pierwszej części przypomniano podstawowe zasady klasyfikacji tematycznej zastosowanej do podziału słownictwa w WSJP PAN, w części drugiej ukazano techniczne sposoby korzystania z wyszukiwarki zaawansowanej słownika do selekcji słownictwa tematycznego, w części trzeciej przedstawiono dane statystyczne dotyczące poszczególnych pól tematycznych w słowniku (według stanu na sierpień 2021 r.) i wstępne wnioski wypływające z ich ogólnej analizy.

## *Wielki słownik języka polskiego PAN* as a thematic dictionary

### Summary

The article describes those features of the *Wielki słownik języka polskiego PAN* (“Polish Academy of Sciences Great Dictionary of Polish”) which make it possible to use it as a thematic

dictionary of Polish (the largest one in terms of the number of the classified units). The first part reminded the basic principles of the thematic classification adopted for the division of the vocabulary, the second part described the technical methods of using the dictionary's advanced search engine to select the thematic vocabularies, the third part presented the statistical data concerning particular thematic fields of the dictionary (valid as of August 2021), and provided the preliminary conclusions following from their general analysis

#### Cytowanie

Żmigrodzki, Piotr. „*Wielki słownik języka polskiego PAN jako słownik tematyczny*”. *Studia Językoznawcze. Synchroniczne i diachroniczne aspekty badań polszczyzny* 21 (2022): 153–166. DOI: 10.18276/sj.2022.21-11.