

Nowe elektroniczne zasoby językowe dla polonistów (słowniki i korpusy)

Słowa kluczowe: humanistyka cyfrowa, słownik, korpus tekstów, językowe zasoby cyfrowe.

<https://doi.org/10.31286/JP.00956>

Wprowadzenie

Polskojęzyczne zasoby humanistyki cyfrowej w ciągu ostatnich kilkunastu lat przyrastają i rozwijają się w bardzo szybkim tempie, o czym świadczą choćby ich zbiorcze omówienia i zestawienia na stronach internetowych typu: *Słowniki dawne i współczesne*¹, *Language tools and resources for Polish*², *CLARIN-PL*³, a także w licznych pracach naukowych, których nie sposób tu wymienić nie tylko ze względu na ich liczbę, ale też ograniczoną aktualność poczynionych w nich ustaleń wynikającą z tempa zachodzących w tym obszarze zmian. W ciągu ostatnich dziesięciu lat próbę zbiorczego przedstawienia dostępnych zasobów elektronicznych podjęli m.in.: Janusz Bień (2014), Milena Hebal-Jeziarska ((red.) 2014), Agata Kwaśnicka-Janowicz i Patrycja Pałka (2015, 2017), Krzysztof Nowak (2015), Magdalena Pastuch, Beata Duda, Karolina Lisczyk, Barbara Mitrenga, Joanna Przyklenk, Katarzyna Sujkowska-Sobisz (zob.: Pastuch i in. 2018; Duda, Lisczyk 2018), Zygmunt Saloni i Włodzimierz Gruszczyński (zob.: Saloni, Gruszczyński 2013; Gruszczyński 2020), Marcin Werla i Maciej Maryl ((oprac.) 2014) oraz Piotr Wierzchoń i Piotr Nowak (2020). Tak duży i zróżnicowany przyrost cyfrowych zasobów językowych z jednej strony znacząco wpływa na rozwój badań naukowych czy metod dydaktycznych, z drugiej zaś strony uniemożliwia badaczom na bieżąco śledzenie nowości oraz zmian zachodzących w już istniejących elektronicznych narzędziach, słownikach, korpusach, bibliotekach bądź kompendiach. Stąd konieczność regularnego aktualizowania zwłaszcza prac zbiorczych, w których omawia się szczegółowo liczne zasoby. Jedną z takich wymagających uzupełnienia i korekty pozycji jest *Przewodnik po elektronicznych zasobach językowych dla polonistów (słowniki, kartoteki, korpusy, kompendia)*, który został wydany w roku 2017 (Pałka, Kwaśnicka-Janowicz 2017). Od tego czasu bowiem część prezentowanych w opracowaniu zasobów przestała być dostępna online (np.: *Słownik wyrazów obcych i zwrotów obcojęzycznych* Władysława Kopalińskiego, *Baza ram interpretacyjnych dla wybranych jednostek czasownikowych* (2009), *Słownik synonimów i antonimów* (2005), a część uległa pewnym

* p.palka@uj.edu.pl; ORCID: 0000-0001-6412-414X

1 Zob. <http://www.leksykografia.uw.edu.pl/> (dostęp: 15 czerwca 2023).

2 Instytut Podstaw Informatyki Polskiej Akademii Nauk, zob. <http://clip.ipipan.waw.pl/LRT> (dostęp: 15 czerwca 2023).

3 Zob. <https://clarin-pl.eu/index.php/zasoby/> (dostęp: 15 czerwca 2023).

modyfikacjom technicznym (np. zmiana adresu strony internetowej) lub merytorycznym (np. dodanie innych możliwości wyszukiwania danych) bądź też po roku 2017 udostępniono całkiem nowe zasoby cyfrowe.

W niniejszym artykule zostaną omówione takie wybrane zasoby elektroniczne, które nie zostały uwzględnione we wspomnianym *Przewodniku po elektronicznych zasobach językowych dla polonistów...* i które wykorzystują materiał współczesnego języka polskiego, a ściślej – opracowują dane językowe datowane od 1945 roku do czasów obecnych lub obejmujące cały wiek XX. Zaznaczmy, że opis poszczególnych leksykonów i korpusów nawiązuje do sposobu przedstawienia zasobów cyfrowych zaproponowanego w *Przewodniku*, dostosowanego do wymienionych poniżej kryteriów ich klasyfikacji. W artykule skupimy się na omówieniu zawartości elektronicznych zasobów językowych (zakresu materiałowego, zasad doboru haseł czy innych danych językowych itp.) oraz technicznych możliwości korzystania z nich. Ewentualne różnice w rozłożeniu akcentów między tymi dwoma aspektami wynikają z odmienności opisywanych źródeł – ich charakteru, struktury oraz przeznaczenia, a także z przyjęcia perspektywy użytkownika. W podsumowaniu omówimy z kolei możliwe sposoby wykorzystania tych zasobów w pracach badawczych i w dydaktyce. Zanim jednak do tego przejdziemy, przypomnijmy zaproponowaną w *Przewodniku* klasyfikację zasobów elektronicznych, w której za nadrzędne kryteria uznano: (1) prymarność/wtórność (lub powtórność) edycji elektronicznej, (2) formę dostępu do zasobów: offline/online oraz (3) stopień możliwości korzystania z danych w zasobach, w tym wyszukiwania danych, czyli: (a) brak możliwości wyszukiwania danych i kopiowania tekstu, (b) ograniczone możliwości korzystania z zasobów (np. wyszukiwanie słów zawierających dany ciąg liter z poziomu okienka dialogowego, przeszukiwanie według listy haseł lub spisu treści), (c) zaawansowane możliwości korzystania z zasobów (np. wyszukiwanie wyrazów i (lub) ciągu znaków w całym słowniku i (lub) w danym haśle, i (lub) w różnych elementach artykułu hasłowego, fakultatywnie: wyszukiwanie haseł w słowniku po wpisaniu dowolnej formy fleksyjnej), (d) bardzo zaawansowane możliwości korzystania z zasobów (np. możliwości wyszukiwania takie jak w punkcie (c), wyszukiwanie określonego ciągu znaków, wszystkich form fleksyjnych danego leksemu, połączeń wyrazowych oraz za pomocą zapytań o klasę gramatyczną, kategorie gramatyczne i metadane typu: autor, tytuł, data wydania itd.) (Pałka, Kwaśnicka-Janowicz 2017: 20–21). I tak w wyniku zastosowania podanych kryteriów P. Pałka i A. Kwaśnicka-Janowicz (2017: 22–23) proponują następującą funkcjonalną klasyfikację zasobów cyfrowych:

1. Zasoby prymarnie wydane elektronicznie (zasoby, które od początku miały postać elektroniczną, czyli „narodziły się jako cyfrowe” [...]):
 - 1.1. Zasoby dostępne offline (nie dostępne za pośrednictwem internetu, wydane na określonych nośnikach danych):
 - 1.1.1. Zasoby bez możliwości wyszukiwania danych (brak przykładu [...]).
 - 1.1.2. Zasoby z ograniczonymi możliwościami korzystania z danych (np. SFJP).
 - 1.1.3. Zasoby z zaawansowanymi możliwościami korzystania z danych (np. SGJP 2012).

- 1.1.4. Zasoby z bardzo zaawansowanymi możliwościami korzystania z danych (brak przykładu [...]).
- 1.2. Zasoby dostępne online (dostępne za pośrednictwem internetu):
 - 1.2.1. Zasoby bez możliwości wyszukiwania danych (brak przykładu [...]).
 - 1.2.2. Zasoby z ograniczonymi możliwościami korzystania z danych (np. ISNWP⁴).
 - 1.2.3. Zasoby z zaawansowanymi możliwościami korzystania z danych (np. WSJP PAN [...]).
 - 1.2.4. Zasoby z bardzo zaawansowanymi możliwościami korzystania z danych (brak przykładu [...]).
2. Zasoby wtórnie lub powtórnie wydane elektronicznie – czyli, po pierwsze, zasoby poddane digitalizacji właściwej, a więc teksty pisane (prymarnie rękopiśmienne lub drukowane) oraz teksty mówione (nagrania zapisane analogowo [...]), a po drugie, zasoby poddane komputeryzacji, czyli teksty prymarnie elektroniczne – „urodzone jako cyfrowe”, już na etapie tworzenia wprowadzone do pamięci komputera, i powtórnie poddane obróbce cyfrowej w celu ich ponownej edycji elektronicznej:
 - 2.1. Zasoby dostępne offline:
 - 2.1.1. Zasoby bez możliwości wyszukiwania danych (brak przykładu [...]).
 - 2.1.2. Zasoby z ograniczonymi możliwościami korzystania z danych (np. BZPPŚ).
 - 2.1.3. Zasoby z zaawansowanymi możliwościami korzystania z danych (np. USJP 2004).
 - 2.1.4. Zasoby z bardzo zaawansowanymi możliwościami korzystania z danych (brak przykładu [...]).
 - 2.2. Zasoby dostępne online:
 - 2.2.1. Zasoby bez możliwości wyszukiwania danych (np. Ksw).
 - 2.2.2. Zasoby z ograniczonymi możliwościami korzystania z danych (np. SGPKał).
 - 2.2.3. Zasoby z zaawansowanymi możliwościami korzystania z danych (np. SJP PWN).
 - 2.2.4. Zasoby z bardzo zaawansowanymi możliwościami korzystania z danych (np. Wyszukiwarka Poliqarp dla NKJP⁵).

Nowe słowniki elektroniczne

Omówione w artykule zasoby językowe, nieuwzględnione w *Przewodniku*, są w większości dostępne online (poza jednym słownikiem), bez konieczności zakładania konta i logowania się; to zarówno prymarne, jak i wtórne czy powtórne wydania cyfrowe. Co więcej, należą one do zasobów z zaawansowanymi bądź bardzo zaawansowanymi możliwościami korzystania z danych, w tym wyszukiwania informacji.

Pierwszą grupę stanowią współczesne leksykony elektroniczne, wśród których dużą wartość mają zasoby reprezentujące prymarnie wydane cyfrowo słowniki onomazjologiczne i tematyczne, mianowicie: słownik onomazjologiczny *Wiktoria* (2009–2011) oraz *Słownik tematyczny WSJP PAN* (ST WSJP PAN, 2021– praca w toku).

4 W cytowanym fragmencie z *Przewodnika* podano w tym miejscu inny przykład, który został wymieniony ze względu na to, że opisane wówczas zasoby obecnie są już niedostępne online.

5 Jw.

*Wiktoria*⁶ ma charakter prototypu słownika onomazjologicznego, gdyż zawiera jedynie 2 tysiące jednostek należących tylko do klasy rzeczowników⁷. Porządek onomazjologiczny (znaczeniowy), jaki przyjęto w słowniku, funkcjonuje na poziomie jego makrostruktury i przedstawiono go na stronie projektu w formie wizualizacji (w postaci koła), gdzie uwagę zwraca antropocentryzm opisu. Centralnym bowiem pojęciem zaproponowanej klasyfikacji onomazjologicznej jest ‘człowiek’, a następnie na niższym poziomie:

‘wiedza o sytuacji w świecie’, ‘wytwory myśli ludzkiej’, ‘świat wartości’, ‘świat otaczający człowieka’, ‘wytwory pracy ludzkich rąk’, ‘zachowanie’, ‘czas wolny’, ‘dziedzictwo kultury’, ‘organizm człowieka i jego funkcjonowanie’, ‘potrawy i napoje’, ‘nazwy osób’, ‘grupa ludzi’, ‘funkcjonowanie człowieka w społeczeństwie’.

Na poziomie mikrostruktury *Wiktoria*, jak podaje się na stronie projektu⁸, „ma być słownikiem ogólnym, uniwersalnym, w którym użytkownik znajdzie wszelkie informacje na temat interesującego go leksemu”, takie jak: definicja, relacje semantyczne, miejsce leksemu w semantycznej strukturze hierarchicznej, informacje gramatyczne, w tym fleksyjne i składniowe. Użytkownik słownika może korzystać z bazy, po pierwsze, przeszukując ją z poziomu głównego okienka dialogowego, w które należy wpisać interesujący nas rzeczownik w mianowniku, ponieważ oprogramowanie nie rozpoznaje form fleksyjnych wyrazów. Po drugie, możemy wyszukiwać dane w sposób zaawansowany według m.in. takich kryteriów, jak: (1) określony ciąg liter (który może rozpoczynać czy kończyć jakąś nazwę bądź się w niej zawierać), (2) miejsce wyszukiwania (forma hasłowa lub odmiana, lub dowolne miejsce w strukturze hasła), (3) typ węzła (leksem albo pojęcie), (4) miejsce w hierarchii semantycznej (liść, rozgałęzienie bądź korzeń), (5) rodzaj gramatyczny oraz (6) relacje semantyczne (typu: metonimia, metafora, holonomia itd.). I wreszcie po trzecie, można przeszukiwać bazę za pomocą dostępnej wizualizacji struktury onomazjologicznej przez przechodzenie od pojęcia nadrzędnego do coraz niższych poziomów (grup pojęciowych), aż do przypisanego końcowo leksemu, np. ‘człowiek’ → ‘czas wolny’ → ‘czas świąteczny’ → *obrzęd, przyjęcie, sakrament, święto, uroczystość*. Oprogramowanie umożliwia także porównywanie wybranych przez użytkownika leksemów.

Z kolei ST WSJP PAN⁹ to największe tego typu dzieło leksykograficzne notujące polszczyzną współczesną (Żmigrodzki 2022b: 132). Uznanie opisywanego zasobu cyfrowego za odrębny słownik wymaga kilku słów wyjaśnienia. Możliwość korzystania z danych językowych WSJP PAN w różnoraki sposób, w tym wyszukiwanie przez użytkownika interesującego go konkretnego zbioru jednostek, np. anglicyzmów, terminów chemicznych, słownictwa młodzieżowego, to właściwość typowa dla słowników wydanych cyfrowo jako „przeszukiwalne bazy danych”

6 <https://wiktoria.ijp.pan.pl/>.

7 Opracowanie tego słownika elektronicznego jest efektem realizacji projektu badawczego pt. *Klasyfikacja onomazjologiczna rzeczowników, a ich charakterystyka gramatyczna* (2009–2011, kierownik grantu: Włodzimierz Gruszczyński, wykonawca: Anna Andrzejczuk).

8 Zob. https://wiktoria.ijp.pan.pl/o_slovníku/ (dostęp: 15 czerwca 2023).

9 https://wsjp.pl/slovník_tematyczny.

i ograniczona jedynie strukturą informatyczną danego zasobu (zob. Żmigrodzki 2022a: 154). W WSJP PAN istnieje zatem także opcja przeszukiwania oraz sortowania materiału leksykograficznego pod kątem jego podziału tematycznego. Nie byłoby to wykonalne (pod względem technicznym), gdyby przed rozpoczęciem opisu tematycznego w WSJP PAN nie opracowano autorskiego schematu klasyfikacyjnego (zob. Batko-Tokarz 2018, 2019; Żmigrodzki 2022b: 134). Dzięki przygotowanej na wstępie, a w dalszej kolejności wprowadzonej konsekwentnie do opisu leksykograficznego haseł w WSJP PAN, klasyfikacji tematycznej polskiej leksyki współczesnej możliwe jest także, z użyciem odpowiedniej funkcji wyszukiwania zaawansowanego, wygenerowanie słownika tematycznego. Co więcej, słownik ów doczekał się już własnego interfejsu, co zapewne wpłynie na większą świadomość wśród użytkowników, że takie opracowanie w ogóle istnieje (zob. Żmigrodzki 2022a: 154).

Omawiany słownik tematyczny, jak już wspomniano, to największy leksykon tego typu – jest on bowiem oparty na ponad 84 tysiącach haseł WSJP PAN (stan z 2022 roku), którego zasób leksykalny stopniowo się rozrasta. A że klasyfikuje się w nim każde znaczenie wyodrębnione w danym artykule hasłowym WSJP PAN, zawiera on ponad 145 tysięcy znaczeń sklasyfikowanych tematycznie, to znaczy przypisanych do określonych sfer, pól i podpól tematycznych, zgodnych z przyjętym podziałem tematycznym słownictwa w języku polskim opracowanym przez Barbarę Batko-Tokarz (Batko-Tokarz 2018; 2019: 135–293; zob. też Żmigrodzki 2022a: 155). Dodajmy tylko, że klasyfikacji tematycznej nie podlegają wyłącznie jednostki funkcyjne, czyli przyimki, spójniki, partykuły itd., zaliczone do sfery – „atematyczne”. Istotne jest to, że stworzona na potrzeby opracowania WSJP PAN klasyfikacja tematyczna ma, podobnie jak w *Wiktorii*, charakter antropocentryczny, co odzwierciedla się zarówno w nazwach poszczególnych pól (czytelnych i zrozumiałych dla użytkowników niebędących specjalistami, np.: „Rodzina”, „Jedzenie i jego przygotowanie”, „Czas wolny”, „Sport”, „Praca”), jak i w nadrzędnej strukturze przyjętego podziału na sfery, która wygląda następująco:

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.

Odnotujmy jeszcze, że opracowany przez B. Batko-Tokarz schemat klasyfikacyjny jest trójstopniowy. Najwyżej w hierarchii znajduje się owych siedem sfer nadrzędnych, w których obrębie wyodrębnia się poszczególne pola (łącznie jest ich 45), a te z kolei dzieli się wewnątrz na szczegółowe podpola (łącznie jest ich 205), np. sfera: II. CZŁOWIEK JAKO ISTOTA PSYCHICZNA → pole 1. Stany psychiczne człowieka → podpola: 1.1. emocje i uczucia, 1.2. zachowania emocjonalne (Batko-Tokarz 2018: 80; 2019: 190). W ST WSJP PAN użytkownik ma możliwość zaawansowanego wyszukiwania haseł w bazie, przypisanych do danej sfery, pola czy podpola. Wyszukiwanie, tak jak i podział tematyczny słownictwa w WSJP PAN, ma charakter trójstopniowy, a więc aby znaleźć odpowiednią kategorię tematyczną i zaklasyfikowane do niej znaczenia, należy w pierwszej kolejności wybrać jedną z ośmiu nadrzędnych sfer. Po kliknięciu w daną sferę otrzymujemy wykaz poszczególnych pól tematycznych reprezentujących

określoną sferę oraz listę wszystkich haseł, których znaczenia do niej zaklasyfikowano. W dalszej kolejności można uszczegółowić wyszukiwanie, klikając w konkretne pole w celu rozwinięcia listy przypisanych do niego podpól tematycznych wraz z wykazem poszczególnych haseł. Oprogramowanie ułatwia odnalezienie interesującego nas znaczenia zaliczonego do określonej sfery i danego pola czy podpola, gdyż w strukturze hasła, będącego jednostką polisemiczną, znaczenie to zaznaczone jest na czerwono.

Wśród nowych zasobów cyfrowych pojawiły się w ciągu ostatnich kilku lat również słowniki ogólny-definicyjne i specjalistyczne – opisujące tylko wybrane typy jednostek leksykalnych, takie jak neologizmy, związki frazeologiczne, antroponimy oraz leksemy pozostające w określonej relacji semantycznej (synonimii, bliskoznaczności, antonimii, hiperonimii itd.) względem innych jednostek. I tak w *Przewodniku* nie uwzględniono wydanego cyfrowo w roku 2018 *Wielkiego słownika języka polskiego PWN ze słownikiem wyrazów bliskoznacznych* (WSJP PWN 2018a; zob. też WSJP PWN 2018b), który dostępny jest offline (możliwe jest uruchomienie słownika z pendrive'a lub zainstalowanie go na wybranym urządzeniu w formie aplikacji). WSJP PWN (2018a) zawiera opis leksykograficzny ponad 100 tysięcy jednostek, wśród których uwzględnia się również wyrazy z najnowszej polszczyzny (ostatniej dekady), frazeologizmy, przysłowia i terminy. Dodatkowo baza ta została uzupełniona o 160 tysięcy wyrazów bliskoznacznych, a także opracowanymi przed Edwarda Polańskiego zasadami pisowni i interpunkcji polskiej. Materiał do WSJP PWN (2018a) zaczerpnięto w dużej mierze z wydanych już (także elektronicznie) zasobów PWN, takich jak: USJP (2003, 2004, 2016) oraz WSWB (2007; zob. też SWB), dlatego też słownik ten należy uznać jednocześnie za wydanie cyfrowe wtórne, gdyż istnieje jego częściowe wydanie drukowane (zob. WSJP PWN 2018b), jak i powtórne, bo oparte jest ono na wcześniejszych publikacjach elektronicznych podanych powyżej słowników. Budowa artykułu hasłowego jest taka sama jak mikrostruktura hasła w USJP (2004, 2016) – po lewej stronie głównego ekranu prezentowane są podstawowe dane leksykograficzne dotyczące danej jednostki leksykalnej (kwalifikator, definicja, przykłady użycia, frazeologizmy czy terminy itp.), a po prawej mamy relacje semantyczne (wyrazy bliskoznaczne oraz sieć powiązań: antonimy, holonimy, hiperonimy, hiponimy, meronimy). Oprogramowanie pozwala na przeszukiwanie słownika na rozmaite sposoby: zarówno z poziomu okienka dialogowego (również z opcją wykorzystania indeksu *a tergo*) po wpisaniu dowolnego ciągu znaków lub wybranej formy fleksyjnej wyrazu, jak i za pomocą narzędzi wyszukiwania zaawansowanego (m.in. z opcją zawężenia zakresu wyszukiwania do samego hasła, przykładów użycia czy grup słownictwa) i krzyżówkowego. Ponadto użytkownik może ograniczyć zakres wyświetlanych haseł do samych przysłów bądź frazeologizmów i terminów, korzystając z tych zbiorów jak z odrębnych słowników elektronicznych. Zaznaczmy, że WSJP PWN to słownik interaktywny, który umożliwia użytkownikowi: dodawanie własnych haseł słownikowych oraz notatek do haseł już istniejących, automatyczne powiązanie pracy z tekstem ze słownikiem (w wyniku czego w szybki, łatwy i bezpośredni sposób można sprawdzić dany wyraz – jego pisownię, znaczenie, użycie czy synonimy itp.), korzystanie z zasobów internetowych online i połączenie się z SJP PWN (2023) oraz z mediami społecznościowymi, profilem PWN na stronie Facebooka.

Nowe jednostki leksykalne notuje z kolei *Słownik neologizmów polskich*¹⁰ (SNP, 2014 – praca w toku), opracowywany w Obserwatorium Językowym Uniwersytetu Warszawskiego kierowanym przez Katarzynę Kłosińską. Słownik ten ma charakter opisowo-rejestrujący, a nie normatywny, i uwzględnia się w nim zarówno nowe wyrazy, związki frazeologiczne, jak i nowe znaczenia leksemów już obecnych we współczesnej polszczyźnie. Co ciekawe, zbiór ten tworzą sami użytkownicy języka, którzy zgłaszają poszczególne jednostki leksykalne za pomocą formularza dostępnego na stronie¹¹. Redakcja weryfikuje zgłoszenia, stosując następujące kryteria: po pierwsze – „jednostka powinna mieć co najmniej 10 poświadczeń w co najmniej 5 powstałych po 2003 r. tekstach przystępnych poznawczo dla przeciętnego użytkownika języka”, a po drugie – „nie może występować w żadnym z ośmiu słowników diagnostycznych wydanych przed 2003 rokiem”¹².

Podstawowa mikrostruktura hasła w SNP obejmuje takie obligatoryjne elementy, jak: (1) wyraz hasłowy, (2) definicja wraz z kwalifikatorami (informującymi o nacechowaniu stylistycznym, ekspresywnym, socjolektalnym i aksjologicznym) oraz (3) przykłady użycia. Fakultatywnie uwzględnia się: (4) warianty wyrazu hasłowego, (5) wymowę, (6) informację gramatyczną, (7) odsyłacze (do haseł powiązanych słotwórczo lub innych, które łączą się z danym wyrazem pod względem semantycznym, pragmatycznym, społecznym czy czasowym) oraz (8) informację graficzną (ilustrację hasła) i (9) informację frekwencyjną. Dodatkowo w artykule hasłowym pojawiają się inne szczegółowe metadane, którymi jednostka została otagowana, dzięki czemu można w zaawansowany sposób przeszukiwać zasoby (zob. sposób tagowania jednostek – lista tagów i symboli z czternastu kategorii, np. typ jednostki, jej rodzimość/obcość, rodzaj derywatu, pochodzenie wyrazu itd.¹³). Oprogramowanie umożliwia zatem wyszukiwanie haseł w bazie z kilku poziomów: (1) wyszukiwarki *a fronte* i (2) wyszukiwarki *a tergo* (zob. zakładka po lewej stronie „Przeczytaj hasła” – należy wybrać z okienka dialogowego dowolny znak, dowolną literę, dowolną cyfrę lub inny znak rozpoczynający/kończący wyraz hasłowy) oraz (3) wyszukiwarki według metadanych (zob. zakładka po lewej stronie „Szukaj haseł”). Ten ostatni sposób przeszukiwania zbioru neologizmów umożliwia wprowadzenie licznych zapytań o cechy danej jednostki leksykalnej, które mają wystąpić lub winny być wykluczone, a są to m.in.: ciągłość czy nieciągłość jednostki, pochodzenie, wartość słotwórcza bądź ekspresywna, wartościowanie, data rejestracji hasła, status jednostki ze względu na kryterium nowości (zob. link do listy tagów).

Z kolei *Verbel*¹⁴ uwzględnia w opisie leksykograficznym jedynie związki frazeologiczne, a dokładnie jest *Słownikiem paradygmatów polskich frazeologizmów czasownikowych* (2018) opracowanym przez Iwonę Kosek, Monikę Czerepowicką oraz Sebastiana Przybyszewskiego.

10 <https://obserwatoriumjezykowe.uw.edu.pl/czym-jest-slownik-neologizmw-polskich/> (dostęp: 15 czerwca 2023).

11 <https://obserwatoriumjezykowe.uw.edu.pl/jak-zglosic-haslo/> (dostęp: 15 czerwca 2023).

12 <https://obserwatoriumjezykowe.uw.edu.pl/czym-jest-slownik-neologizmw-polskich/> (dostęp: 15 czerwca 2023).

13 <https://obserwatoriumjezykowe.uw.edu.pl/zasady-opisu-hasel/> (dostęp: 15 czerwca 2023).

14 <http://uwm.edu.pl/verbel/>. Wprawdzie możliwy jest pełny dostęp do słownika bez konieczności założenia konta i logowania się, ale aby korzystać z bazy, należy uprzednio pobrać dane lingwistyczne oraz zainstalować program Werbosław (zgodnie z instrukcją – zob. <http://uwm.edu.pl/verbel/wp-content/uploads/2018/10/Instalacja-programu-1.pdf>, dostęp: 15 czerwca 2023).

W bazie notuje się i opisuje ponad 5 tysięcy zwrotów (w formie trzecioosobowej typu *ktoś/coś postawi coś na głowie*), które zostały wyekscerpowane z różnych źródeł (zob. Kosek i in. 2018), w szczególności z NKJP oraz z wybranych słowników ogólnych języka polskiego (np.: SWJP, USJP, WSJP PAN) i ze słowników frazeologicznych (np. SFWP). W opracowaniu uwzględnia się przede wszystkim formy fleksyjne frazeologicznych jednostek czasownikowych, ale także informację o znaczeniu i aspekcie, odpowiedniku w parze aspektowej (jeśli takowa istnieje) oraz ewentualne dodatkowe komentarze o użyciu danego zwrotu – jego pragmatyce, zakresie czy frekwencji. Ponadto, jak wyjaśniają I. Kosek i S. Przybyszewski, w słowniku podawane są też informacje dotyczące „gramatyki poszczególnych komponentów, [...] wartości kategorii liczby (i ewentualnie rodzaju) składników nominalnych”, a więc tego, co odnosi się „do składni wewnętrznej związku frazeologicznego (w rozumieniu Lewickiego [...]) i możliwości korelacji postaci zwrotu z kontekstem zewnętrznym” (zob. „Wprowadzenie do słownika”¹⁵). Na mikrostrukturę hasła składają się zatem takie elementy opisu leksykograficznego, jak: (1) opis ogólny hasła (postać zwrotu, aspekt, możliwe znaczenia i inne komentarze), (2) opis jednostki (budowa wewnętrzna zwrotu – opis gramatyczny jego członów, w tym określenie członu głównego), (3) opis odmiany fleksywu¹⁶ (zestaw metadanych), (4) formy fleksywu, (5) wszystkie formy fleksyjne jednostki (zob. Kosek i in. 2020). Hasła słownikowe można zaś wyszukiwać na dwa sposoby: przeglądając alfabetyczną listę opisanych frazeologicznych jednostek czasownikowych lub korzystając z okienka dialogowego, w którym należy wpisać wybrany komponent jakiegoś zwrotu. Program umożliwia dodatkowo wiele operacji, w tym m.in. ustawienie filtrów grafów, jednak korzystanie z niego nie jest do końca intuicyjne, wymaga specjalistycznej wiedzy i głębszego poznania mechanizmów technicznych. W łatwy sposób można jedynie przeglądać hasła według listy alfabetycznej i wyszukiwać je z poziomu okienka dialogowego.

Jeśli zaś chodzi o zasoby z zakresu onomastycznej leksykografii cyfrowej, to powiększyły się one o *Internetowy słownik nazwisk w Polsce*¹⁷ (ISNwP 2023). Ten prymarnie wydany elektronicznie onomastykon opracowywany jest od 2014 roku w Instytucie Języka Polskiego PAN. Powstał z inicjatywy Barbary Czopek-Kopciuch, a jego autorkami są: Halszka Górny, Małgorzata Magda-Czekaj, Elena Palinciuc-Dudek, Katarzyna Skowronek oraz Elżbieta Supranowicz. Słownik notuje około 30 tysięcy nazwisk współcześnie używanych w Polsce, pozyskanych z rejestru bazy danych PESEL z 2014 roku. Dodatkowymi źródłami odnoszącymi się do danych historycznych są odpowiednie naukowe opracowania i słowniki onomastyczne¹⁸. Na maksymalną mikrostrukturę artykułu hasłowego składają się: (1) forma główna nazwiska będąca tytułem hasła, (2) formy żeńskie nazwiska, (3) informacje o frekwencji – liczbie użytkowników danego nazwiska i jego randze w Polsce, (4) geografia nazwiska w Polsce, (5) zapisy historyczne, (6) pochodzenie, prawdopodobne etymologie oraz motywacje nazwiska, (7) opis

15 <http://uwm.edu.pl/verbel/wp-content/uploads/2018/10/wprowadzenie-1.pdf> (dostęp: 15 czerwca 2023).

16 *Fleksywu*, jak wyjaśniają autorzy słownika, rozumiany jest jako „zbiór form jednolicie lub niemal jednolicie zróżnicowanych ze względu na właściwe im kategorie gramatyczne. Traktowanie jako jednostek opisu fleksywów, nie leksemów, niesie z sobą konsekwencje zarówno teoretyczne, jak i techniczne” (Czerepowicka i in. 2014: 120).

17 <https://nazwiska.ijp.pan.pl/> (dostęp: 15 czerwca 2023).

18 Zob. opis źródeł: <https://nazwiska.ijp.pan.pl/info/oslowniku> (dostęp: 15 czerwca 2023).

słowotwórczy nazwiska, (8) gwarowe, fonetyczne i graficzne formy wariantywne nazwiska, (9) frekwencja i rozmieszczenie nazwisk w poszczególnych krajach, (10) informacje dodatkowe, (11) linki do zagranicznych cyfrowych onomastykonów notujących imiona i nazwiska, (12) literatura przedmiotu. Tak rozległy sposób opisu jednostek onimicznych sprawia, że mamy do czynienia nie tylko ze słownikiem, ale też z kompendium antroponicznym (zob. Skowronek i in. 2018: 301). Co ważne, aplikacja umożliwia wyświetlenie wersji do druku oraz różne sposoby wyszukiwania danych. Można bowiem zarówno przeglądać alfabetyczną listę haseł, jak i skorzystać z okienka dialogowego, gdzie wpisuje się formę poszukiwanego nazwiska bądź ciąg liter, który rozpoczyna, kończy lub zawiera dane hasło. Odnotujmy także, że ISNWP nie jest dziełem skończonym, prace nad nim są kontynuowane, a baza jest rozwijana i uzupełniana.

Nowe elektroniczne korpusy języka polskiego

Istotnym *novum* w polskiej lingwistyce komputerowej jest rozwój zasobów, które dają łatwy dostęp do ogromnych, lematyzowanych i anotowanych morfoskładniowo oraz za pomocą różnych metadanych, zbiorów tekstów reprezentujących dany typ dyskursu czy społeczną odmianę języka, umożliwiając ich przeszukiwanie i filtrowanie z użyciem bardzo zaawansowanych narzędzi informatycznych. Mamy tu na myśli zwłaszcza dwa przykłady najnowszych zasobów cyfrowych: *Korpus dyskursu parlamentarnego*¹⁹ (KDP 2022) oraz *ChronoPress – korpus tekstów prasowych*²⁰ (ChP 2023).

KDP został opracowany przez Zespół Inżynierii Lingwistycznej IPI PAN działający pod kierownictwem Macieja Ogrodniczuka. Baza zawiera anotowane lingwistycznie (według tych samych zasad co NKJP) teksty dokumentujące prace obu izb parlamentarnych II i III RP oraz Sejmu w dobie PRL-u²¹ od roku 1919 do czasów współczesnych. Zgodnie z metadanymi, jakie zostały przypisane dokumentom, w bazie uwzględnia się następujące typy tekstów dyskursu parlamentarnego: (1) posiedzenie plenarne, (2) komisja, (3) zgromadzenie posłów i senatorów, (4) wspólne posiedzenie Sejmu i Senatu, (5) Zgromadzenie Narodowe i (6) uroczyste spotkanie, a także (7) uroczyste zgromadzenie posłów i senatorów, (8) interpelacja, (9) odpowiedź na interpelację oraz (10) zapytanie i (11) odpowiedź na zapytanie. Zbiór źródeł jest stale uzupełniany o nowy materiał językowy, a obecna wielkość korpusu to 800 milionów segmentów²². System umożliwia przeszukiwanie korpusu na dwa sposoby. Po pierwsze, można skorzystać z wyszukiwarki segmentów²³, w której zapytanie formułuje się samemu za pomocą odpowiedniej składni zapytań²⁴ (zbliżonej do składni zapytań w PoliQarpie w NKJP). Możemy

19 https://kdp.nlp.ipipan.waw.pl/query_corpus/. Inne wersje elektroniczne dostępne online – korpus do pobrania w postaci plików (format XML TEI P5) ze strony: <http://clip.ipipan.waw.pl/PPC> (dostęp: 15 czerwca 2023).

20 <https://chronopress.clarin-pl.eu/> (dostęp: 15 czerwca 2023).

21 Jak podaje się na stronie KDP: „Wszystkie teksty pochodzą z serwisów Sejmu i Senatu RP i zostały udostępnione dzięki uprzejmości Kancelarii Sejmu RP, Kancelarii Senatu RP oraz Biblioteki Sejmowej. Teksty stanowią informację publiczną i jako takie są dostępne w domenie publicznej” (<https://kdp.nlp.ipipan.waw.pl/overview>, dostęp: 15 czerwca 2023).

22 Szczegółowy opis stanu danych korpusowych: <http://clip.ipipan.waw.pl/PPC> (dostęp: 15 czerwca 2023) (zob. też: Ogrodniczuk, Nitoń 2020; Ogrodniczuk i in. 2022).

23 Zob. https://kdp.nlp.ipipan.waw.pl/query_corpus/ (dostęp: 15 czerwca 2023).

24 Zob. <https://kdp.nlp.ipipan.waw.pl/manual#zapytania> (dostęp: 15 czerwca 2023).

wyszukać np.: (a) wszystkie formy fleksyjne danego leksemu, jak choćby rzeczownika *ojczyzna*, wpisując [base="ojczyzna"], (b) konkretną formę fleksyjną danego leksemu, wpisując ją bezpośrednio w okienko dialogowe: *Polską* lub stosując odpowiednią składnię: [orth="Polską"], (c) dane słowo niezależnie od jego zapisu w tekstach wielką czy małą literą: [orth_lc="polski"], (d) określone klasy i kategorie gramatyczne – zgodnie z przyjętymi skrótami w języku angielskim²⁵ typu: [pos="subst"] lub za pomocą umieszczonego pod okienkiem dialogowym konstruktora zapytań umożliwiającego wprowadzenie określonych atrybutów segmentu/segmentów, takich jak: część mowy, liczba, rodzaj, aspekt itd. System umożliwia także wyszukiwanie segmentów z zawężeniem wyników zgodnie z dokonanym wyborem określonych metadanych, takich jak: izba, typ tekstu, mówca, data, rok, ustrój, numer kadencji, sesja oraz etykieta i tytuł. Ponadto dodatkowym narzędziem ułatwiającym grupowanie i wizualizację wyników wyszukiwania jest panel „Statystyki”, za pomocą którego można na przykład uzyskać szczegółowe dane ilościowe dotyczące występowania danego wyrazu w dyskursie parlamentarnym z podziałem na poszczególne lata (innymi kryteriami, ze względu na które można grupować dane, są: etykieta, tytuł, kadencja, sesja, dzień, forma tekstowa, lemat, część mowy). Po drugie, KDP umożliwia skorzystanie z wyszukiwarki tekstów, jeśli poszukujemy konkretnych materiałów językowych i całych dokumentów ograniczonych zgodnie z zaznaczonymi metadanymi typu: izba, typ tekstu, mówca, data, rok, ustrój, numer kadencji, sesja oraz etykieta i tytuł. Następnie interesujące nas dokumenty możemy pobrać w postaci pliku PDF lub ZIP.

Z kolei *ChronoPress* zawiera około 100 tysięcy fragmentów tekstów prasowych z lat 1945–1964. W dalszej perspektywie planuje się uwzględnienie w bazie również tekstów wcześniejszych i późniejszych, tak aby obejmowały one stulecie polskiej niepodległości (1918–2018). Teksty są oznakowane morfosyntaktycznie i za pomocą metadanych. ChP został opracowany przez zespół współpracowników z Uniwersytetu Wrocławskiego i Politechniki Wrocławskiej pod kierunkiem Adama Pawłowskiego oraz w ramach działań konsorcjum naukowego CLARIN-PL. Baza ta, jak wyjaśnia się na stronie korpusu²⁶, „daje wgląd w dyskurs prasowy PRL” i „oferuje [...] rzetelny i przekrojowy przegląd tematów prasy codziennej tamtych lat – lat, które ukształtowały obraz świata i wrażliwość komunikacyjną kilku pokoleń Polaków”. Warto zaznaczyć, że ChP jest przykładem nie tylko korpusu specjalistycznego, gromadzącego teksty reprezentujące daną specjalistyczną odmianę języka (tu zarówno dyskursu medialnego – prasowego, jak i publicznego czy politycznego), ale też i korpusu chronologicznego, który, zdaniem A. Pawłowskiego (2016: 60; 2017: 2), nie powinien być utożsamiany z pojęciem korpusu diachronicznego: „Korpus taki [chronologiczny – P.P.] charakteryzuje się ścisłym i wyrażonym za pomocą metadanych uporządkowaniem sekwencyjnym tworzących go tekstów. Dzięki temu możliwe jest tworzenie rozłożonych na osi czasu histogramów częstości leksemów lub innych jednostek”. W korpusie chronologicznym próbki tekstów posegregowane są „na osi czasu według dat wydania lub powstania” i są one jednolite „pod względem ortograficznym i typograficznym” (Pawłowski 2016: 61). Tak więc przedmiotem badań w korpusie chronologicznym, jak

25 Zob. <http://nkjp.pl/poliquarp/help/ense2.html#x3-30002.1> (dostęp: 15 czerwca 2023).

26 Zob. zakładka „Prasa w PRL” na stronie: <https://chronopress.clarin-pl.eu/#/> (dostęp: 15 czerwca 2023).

o tym pisze A. Pawłowski, jest przede wszystkim stała forma wyrazowa (jednostka niezmienna pod względem budowy czy zapisu), a w korpusie diachronicznym – jednostka podlegająca zmianom zachodzącym w systemie językowym (Pawłowski 2016: 60; 2017: 2). Anotacja ChP umożliwia przeszukiwanie korpusu według rozmaitych kryteriów i na wiele sposobów za pomocą zaprojektowanych sześciu konstruktorów zapytań. Po pierwsze, można pozyskać dane na temat częstości użycia wyrazów zarówno w postaci list frekwencyjnych²⁷, jak i przez wygenerowanie szeregów czasowych dla określonego słowa, informujących o ewolucji frekwencji tegoż słowa w poszczególnych latach lub miesiącach²⁸. Po drugie, oprogramowanie pozwala na przeglądanie korpusu pod względem występujących w nim częstych połączeń wyrazowych, a ściślej – umożliwia określenie profili kolokacyjnych wyrazów²⁹ oraz wizualne porównanie dwóch leksemów pod kątem ich kolokatów (wykazanie różnic i podobieństw w zakresie występowania wybranych dwóch leksemów w określonym kontekście słownym³⁰). Po trzecie, w korpusie ChP możliwe jest przeglądanie większych próbek tekstów³¹ oraz konkordancji³² po wpisaniu w okienko dialogowe danego leksemu (formy hasłowej) bądź wyrazu w interesującej nas formie odmienionej. I wreszcie po czwarte, we wszystkich opisanych powyżej sposobach wyszukiwania danych w korpusie można ograniczyć wyniki przez wybranie określonych filtrów, takich jak: rok publikacji, tytuł periodyku, autorzy, środek przekazu, ekspozycja (pierwsza strona, środek, ostatnia strona), okresowość (dziennik, tygodnik, miesięcznik itp.) oraz status obiegu i styl (zob. Pawłowski 2021). Dodajmy jeszcze na koniec, że narzędzia korpusowe ChP umożliwiają także przypisanie rozpoznanych w bazie toponimów do punktów na mapie (Google maps) (zob. Pawłowski 2017: 3–4).

Wybrany typem tekstów zajęli się także twórcy *Mikrokorpusu metafor synestezyjnych* (2019), funkcjonującego pod nazwą SYNAMET³³. Jest to baza niewielka ilościowo, ale znacząca pod względem jakościowym i metodologicznym. Korpus zawiera bowiem 9217 jednostek metaforycznych (opisanych gramatycznie i semantycznie za pomocą ram interpretacyjnych w ujęciu Charlesa J. Fillmore’a), które w bazie (zgodnie z przyjętą koncepcją opisu) są reprezentowane nie tylko przez pojedyncze wyrazy, ale też i frazy, a nawet całe zdania czy fragmenty tekstu. SYNAMET jako pierwszy korpus metafor w Polsce, a jedyny korpus metafor synestezyjnych³⁴

27 Zob. https://chronopress.clarin-pl.eu/#/frequency_lists (dostęp: 15 czerwca 2023).

28 Zob. https://chronopress.clarin-pl.eu/#/time_series (dostęp: 15 czerwca 2023).

29 Zob. https://chronopress.clarin-pl.eu/#/word_profiles (dostęp: 15 czerwca 2023).

30 Zob. <https://chronopress.clarin-pl.eu/#/similarity> (dostęp: 15 czerwca 2023).

31 Zob. <https://chronopress.clarin-pl.eu/#/samples> (dostęp: 15 czerwca 2023).

32 Zob. <https://chronopress.clarin-pl.eu/#/concordance> (dostęp: 15 czerwca 2023).

33 <http://synamet.polon.uw.edu.pl/korpus> (dostęp: 15 czerwca 2023). Dodatkowe możliwości oprogramowania (pliki do pobrania i wykorzystania w niekomercyjnych celach naukowych): program ATOS – narzędzie do gramatycznej i semantycznej anotacji metafor oraz ontologia ramowa.

34 Jak wyjaśnia się na stronie projektu, prace dotyczyły „ograniczonego obszaru badawczego (metafor synestezyjnych) i skupiały się na analizie autentycznych, niepreparowanych tekstów z blogów. Metafora synestezyjna polega m.in. na użyciu jednej nazwy do opisu rozmaitych wrażeń zmysłowych, np. *słodki zapach/kolor/dźwięk*. Metaforę synestezyjną w niektórych ujęciach rozszerza się na wszystkie wyrażenia, w których występują wyrazy odnoszące się prymarnie do jakiegoś typu postrzegania, czyli będą to zarówno frazy, typu *jasny dźwięk* czy *słodki zapach*, jak [i] konstrukcje *jasny umysł* lub *ciemna sprawa*” (zakładka „O projekcie”, <http://synamet.uw.edu.pl/>, dostęp: 15 czerwca 2023).

na świecie, został opracowany przez zespół badaczy pod kierunkiem Magdaleny Zawisławskiej z Uniwersytetu Warszawskiego. Materiał źródłowy do korpusu został zaczerpnięty z 1414 tekstów blogów, których autorzy recenzowali takie produkty, jak piwa, wina, kawy, perfumy oraz muzyka. W takich bowiem tekstach była, zdaniem autorów korpusu, największa szansa na zależenie metafor synestezyjnych (Zawisławska 2019). Korpus można przeszukiwać przez wpisanie w okienko dialogowe konkretnego wyrazu, który występuje w jednostkach metaforycznych i aktywuje określoną ramę semantyczną³⁵. Ponadto oprogramowanie pozwala na wyszukiwanie metafor z ograniczeniem wyboru co do ich cech gramatycznych (np.: grupa rzeczownikowa, grupa liczebnikowa, grupa przymiotnikowa, grupa przyimkowo-nominalna, grupa werbalna prosta lub złożona) bądź semantycznych (np.: określona rama źródłowa czy docelowa, konkretny typ metafory złożonej metafory).

I wreszcie ostatnim z wybranych do opisu korpusów nieuwzględnionych w wydaniu *Przewodnika* z 2017 roku jest *Narodowy fotokorpus języka polskiego*³⁶ (NFJP 2014–2020). Baza zawiera, po pierwsze, udokumentowane w postaci fotocytatów słownictwo polskie ekscerpowane z tekstów źródłowych, w tym także w dużej mierze prasowych, z lat 1901–2000, składające się na 250 tysięcy haseł słownikowych. Zgodnie z przyjętymi kryteriami do zbioru tego winien był trafić każdy leksem nieodnotowany w pierwszym XX-wiecznym słowniku ortograficznym autorstwa Władysława Kokowskiego (1903) (Wawrzyńczyk i in. 2021: 10). Po drugie, projekt, kierowany do roku 2021 przez Jana Wawrzyńczyka, ma charakter rozwojowy i obejmuje także inne okresy. W korpusie uwzględnia się bowiem takie odrębne segmenty konstrukcji fotokorpusu, jak lata: 1773–1800, 1801–1900 i 2001–2100. Główni inicjatorzy i twórcy NFJP podają jednak, że „w odpowiednich miejscach strony internetowej udostępniamy z tych lat jedynie nieliczne, przypadkowe fotocytaty; poważna rozbudowa tych segmentów to sprawa kilku najbliższych lat” (Wawrzyńczyk i in. 2021: 12). Na dany artykuł hasłowy składa się zawsze wraz z danymi bibliograficznymi poświadczenie fotocytatowe (zwykle zdaniowe, czasem wielozdaniowe) danej jednostki leksykalnej, będące, co wyjaśnia J. Wawrzyńczyk, udokumentowanym ekscerptem „w postaci fotodokumentacyjnej, tzn. w takiej, w jakiej ukazał się on w druku nim wszedł w nasze e-posiadanie”³⁷. Taki sposób podawania informacji o użyciu leksemu gwarantuje, zdaniem autorów NFJP, „absolutną zgodność wyimka z jego źródłem drukowanym (ewentualnie, zupełnie wyjątkowo – rękopiśmiennym)” (Wawrzyńczyk i in. 2021: 9). W strukturze hasła zamieszczone są ponadto takie elementy opisu leksykograficznego, jak: etykieta gramatyczna poświadczenia (np. informacja o części mowy, liczbie), informacje dodatkowe w postaci wykresu dotyczące diachronicznej częstości użycia słowa (wystąpień na milion wyrazów), lokalizacja ekscerptu na stronie oraz alfabetyczna lista (w układzie *a fronte* i *a tergo*) wyrazów sąsiadujących z interesującym nas leksemem w spisie haseł. Na liście tej znajdują się również leksemu opatrzone symbolem „K”, świadczącym o tym, że jest to jednostka

35 Mowa tu o *aktywatorze* – wyrazie lub wyrażeniu aktywującym „ramę pojęciową”, np. w wyrażeniu *hard rock szyty prawdziwą bluesową miarą* wyrazy *hard rock*, *bluesowy* aktywują ramę SŁUCH, natomiast wyrazy *szyć*, *miara* aktywują ramę UBRANIE” (<http://synamet.polon.uw.edu.pl/korpus>, dostęp: 15 czerwca 2023).

36 <https://nfjp.pl> (dostęp: 15 czerwca 2023).

37 Opis projektu na stronie głównej: <https://nfjp.pl/> (dostęp: 15 czerwca 2023).

nieodnotowana w NFJP, gdyż obecna w słowniku W. Kokowskiego (1903). Jeśli zaś chodzi o możliwości wyszukiwania danych w korpusie, to w pierwszej kolejności możemy przeszukiwać bazę za pomocą indeksu w postaci alfabetycznej listy haseł³⁸. Prezentacja materiału w układzie alfabetycznym wygląda następująco: pod każdą literą widoczny jest zbiór jednostek leksykalnych rozpoczynających się od danej litery wraz z ich łączną liczbą w korpusie. Dodatkowo po kliknięciu na dany zbiór (na daną literę) wyświetla się wizualna informacja diachroniczna (w formie osi czasu) na temat częstości użycia wybranych do oglądu wyników. W dalszej kolejności możemy wyszukiwać jednostki leksykalne z poziomu okienka dialogowego znajdującego się w prawym górnym rogu strony. System nie rozpoznaje form fleksyjnych wyrazów, lecz wyszukuje hasła tylko według ciągu znaków, jaki zawiera dana jednostka. Jeśli więc wpisujemy słowo *propaganda*, wyniki będą następujące: brak hasła *propaganda* w NFJP, ale obecne hasła takie jak: *antypropaganda*, *hurrapropaganda*, *kontrpropaganda* itp. Jeśli zaś wpisujemy ciąg liter *propago*, system pokaże nam fotocyfaty dla jednostek typu: *niepropagowanie*, *propagowanie*, *rozpropagowanie*, *rozpropagować* itp.

Podsumowanie

Przedstawione w niniejszym tekście zasoby cyfrowe otwierają wiele możliwości i perspektyw w zakresie badań językoznawczych, socjologicznych, kulturoznawczych, antropologicznych czy historycznych, a także pod kątem wykorzystania ich w dydaktyce, o czym szczegółowo pisano już na przykładzie zarówno językoznawczej dydaktyki akademickiej, jak i nauczania języka polskiego jako drugiego oraz w szkole (zob.: Duda, Lisczyk 2018; Kwaśnicka-Janowicz, Pałka 2021). Podane zasoby umożliwiają bowiem szybki, łatwy, i co dziś ważne dla użytkownika zwłaszcza nieprofesjonalnego, przyjazny dostęp, po pierwsze, do interesującego daną osobę słownictwa tematycznego należącego do wybranego pola czy podpola (zob. ST WSJP PAN), po drugie, do opracowanej przez językoznawców, a nie amatorów, najnowszej polszczyzny – niewuwzględnionej jeszcze w słownikach cyfrowych, a tym bardziej w tradycyjnych słownikach drukowanych, obecnej na przykład w slangu młodzieżowym, dyskursie reklamowym czy internetowym (zob. SNP). I w końcu po trzecie, wyróżnione zasoby, a konkretnie ISNWB, odpowiadają na współczesne potrzeby społeczne w zakresie konstruowania osobistej i rodzinnej tożsamości oraz

podtrzymywania własnej autonarracji biograficznej [...] przez ułatwianie wszelkiego typu poszukiwań genealogicznych i rodzinnych oraz dostarczanie wiedzy naukowej o nazwisku; przez utrwalanie wiedzy historycznej o danej formie nazwiskowej (Skowronek i in. 2018: 300, 302).

W zakresie zastosowania przedstawionych cyfrowych słowników i korpusów w badaniach językoznawczych można wyróżnić kilka aspektów i naukowych kręgów zainteresowań. Pierwszym z nich są badania z zakresu leksykologii, które mogą ściśle dotyczyć wyłącznie

38 Zob. <https://nfjp.pl/index/> (dostęp: 15 czerwca 2023).

semantyki, w tym tematycznego podziału współczesnej leksyki polskiej (z ograniczeniem się do analizy wybranej czy wybranych grup tematycznych; zob. ST WSJP PAN), analizy pól pojęciowo-leksykalnych (zob.: ST WSJP PAN, *Wiktoria*), profili kolokacyjnych, słów kluczowych (zob.: ChP, KDP) oraz takich zjawisk językowych, jak wartościowanie, metaforyzacja bądź prozodia semantyczna (zob.: ChP, KDP, SNP, SYNAMET). Analiza leksykologiczna może również wiązać się z wyborem do opisu semantycznego lub formalnego określonych typów neologizmów – zapożyczeń, zestawień, derywatów czy frazeologizmów (zob.: NFJP, SNP, WSJP PAN). Współczesne korpusy językowe umożliwiają ponadto szczegółowe badania nad słownictwem polskim pod kątem frekwencji wyrazów (możliwe jest generowanie list frekwencyjnych, np. w ChP) oraz szeregów czasowych dla danych leksemów, a więc przeprowadzenie analizy chronologicznej, która ukazuje „w formie wykresów ewolucję frekwencji wybranych leksemów w czasie”³⁹. Drugim z możliwych kontekstów wykorzystania omówionych zasobów cyfrowych są badania nad gramatyką języka polskiego, a ściśle – nad własnościami składniowymi lub fleksyjnymi wybranych jednostek leksykalnych, w tym także związków frazeologicznych, które można analizować za pomocą zaawansowanych narzędzi korpusowych (np.: ChP, KDP) lub korzystając z już zaproponowanego opisu języka i określonej metodologii językoznawczej (np.: *Verbel*, SYNAMET). Trzeci krąg możliwych do rozwijania dzięki przedstawionym słownikom i korpusom zainteresowań badawczych to onomastyka. Omówiony bowiem ISNWP może stać się ważną podstawą różnorodnych syntez z zakresu antroponimii polskiej w ujęciu historycznym i typologicznym. Z kolei korpusy ChP i KDP umożliwiają wyszukiwanie i analizę użyć w dyskursie prasowym i parlamentarnym nie tylko antroponimów, ale też i innych jednostek onimicznych. Również pragmalingwistyka, tekstologia oraz analiza dyskursu to dziedziny badawcze, w których można analizować, zwłaszcza za pomocą opisanych korpusów cyfrowych KDP i ChP, m.in. takie zjawiska językowe, jak: grzeczność językowa (np. stosowane formy adresatywne) oraz agresja językowa, słowa kluczowe (ważne dla danego typu dyskursu, jego wartości, kreowanej ideologii i wizji świata), wyznaczniki stylistyczne i kształtowanie się stylu w jego warstwie leksykalnej i składniowej, język nowomowy czasów PRL-u, a także modelowanie tematyczne tekstów (wydźwięk emocjonalny tekstów oraz wątki tematyczne i kontekstowe) i wiele innych. Warto jeszcze wspomnieć, że bardzo ciekawe perspektywy badawcze otwierają się przed językoznawcami w zakresie możliwych analiz porównawczych z uwzględnieniem w korpusie zwłaszcza kryterium czasu.

Tak szybko i prężnie rozwijająca się współcześnie humanistyka, w tym lingwistyka cyfrowa, umożliwia prowadzenie bardzo urozmaiconych, rzetelnych, opartych na dużych próbkach tekstów prac naukowych i działań dydaktycznych. Ze względu jednak na liczbę elektronicznych zasobów językowych zachodzi konieczność nie tylko ciągłej aktualizacji opracowań zbiorczych, ale też zapewnienia dostępu do platform, które będą integrowały funkcjonujące już cyfrowe słowniki i korpusy. Takim wstępnym projektem jest *Platforma leksykalna*⁴⁰ powstała w ramach infrastruktury CLARIN-PL, służąca właśnie do integracji różnych elektronicznych zasobów

39 <https://clarin-pl.eu/index.php/chronopress/> (zob. też Pawłowski 2016).

40 <http://lexp.clarin-pl.eu/> (dostęp: 15 czerwca 2023).

leksykalnych, takich jak: *Słowsieć*, *Walenty*, SGJP 2020, HASK, SPXVI, ESPXVII, eSwil, a także Open Multilingual WordNet (zob. Piasecki i in. 2018). Drugą tego typu inicjatywą CLARIN-PL jest *Słownik gramatyczny jednostek wielowyrazowych*⁴¹, nad którym obecnie trwają prace i który ma być dostępny online jako baza scalająca już opracowane zasoby językowe: SAWA, SEJF, SEJFEK i *Verbel*. Miejmy nadzieję, że projekty te będą w najbliższych latach rozwijane i udoskonalane lub też powstanie nowa baza, która umożliwi badaczom oraz dydaktykom korzystanie ze scalonych cyfrowych zasobów językowych diachronicznych i synchronicznych.

Bibliografia

- Batko-Tokarz B. 2018: *Klasyfikacja tematyczna w WSJP PAN*, [w:] P. Żmigrodzki, M. Bańko, B. Batko-Tokarz, J. Bobrowski, A. Czelakowska, M. Grochowski, R. Przybylska, J. Waniakowa, K. Węgrzynek (red.), *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*, Instytut Języka Polskiego Polskiej Akademii Nauk, Kraków.
- Batko-Tokarz B. 2019: *Tematyczny podział słownictwa współczesnego języka polskiego. Teoria, praktyka, leksykografia*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.
- Baza ram interpretacyjnych dla wybranych jednostek czasownikowych, 2009 (online: <http://www.ramki.uw.edu.pl/>, <http://www.ramki.uw.edu.pl/ramki/index.html>).
- Bień J.S. 2014: *Humanistyka cyfrowa. Wybrane przykłady*, prezentacja podczas konferencji „Humanistyka cyfrowa. Projekty, ludzie, teoria”, Warszawa, 3 listopada 2014 (<https://core.ac.uk/download/pdf/77128838.pdf>, dostęp: 15 czerwca 2023).
- BZPPŚ: *Biblioteka zabytków polskiego piśmiennictwa średniowiecznego*, red. W. Twardzik, Instytut Języka Polskiego Polskiej Akademii Nauk, Kraków 2006 (DVD).
- ChP: A. Pawłowski, *ChronoPress – Korpus Tekstów Prasowych*, Repozytorium CLARIN-PL, 2021 (<https://chronopress.clarin-pl.eu/#/>, dostęp: 15 czerwca 2023).
- CLARIN-PL (<https://clarin-pl.eu/index.php/zasoby/>, dostęp: 15 czerwca 2023).
- Czerepowicka M., Kosek I., Przybyszewski S. 2014: *O projekcie elektronicznego słownika odmiany frazeologizmów czasownikowych*, „Polonica” XXXIV, s. 115–123.
- Duda B., Liszczyk K. 2018: *Narzędzia cyfrowe w polonistycznej dydaktyce akademickiej – zastosowania, możliwości, perspektywy*, „Forum Lingwistyczne”, nr 5, s. 143–154.
- ESJXXVII: *Elektroniczny słownik języka polskiego XVII i XVIII wieku*, red. W. Gruszczyński, Instytut Języka Polskiego Polskiej Akademii Nauk (<https://sxvii.pl/>).
- eSwil: Elektroniczna edycja tzw. Słownika wileńskiego: *Słownik języka polskiego*, red. A. Zdanowicz i in., staraniem i kosztem M. Orgelbranda, t. 1–2, Wilno 1861 (<http://eswil.ijp.pan.pl/>).
- Gruszczyński W. 2020: *Korpusy językowe narzędziem pracy historyka języka*, [w:] A. Hącia, K. Kłosińska, P. Zbróg (red.), *Polszczyzna w dobie cyfryzacji*, Polska Akademia Nauk, Warszawa, s. 255–265.
- HASK: Interaktywna baza HASK, 2011–2012 (http://pelcra.pl/hask_pl/, dostęp: 15 czerwca 2023).
- Hebal-Jeziarska M. (red.) 2014: *Praktyczny przewodnik po korpusach języków słowiańskich*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa (<http://slawistyka.uw.edu.pl/wp-content/uploads/sites/283/2014/11/Praktyczny-przewodnik-po-korpusach-j%C4%99zyk%C3%B3w-s%C5%82owia%C5%84skich.pdf>, dostęp: 5 maja 2023).
- ISNWP: *Internetowy słownik nazwisk w Polsce*, 2023 (<https://nazwiska.ijp.pan.pl/>, dostęp: 15 czerwca 2023).
- KDP: *Korpus dyskursu parlamentarnego*, 2022 (https://kdp.nlp.ipipan.waw.pl/query_corpus/, dostęp: 15 czerwca 2023).
- Kokowski W. 1903: *Słownik ortograficzny języka polskiego*, Nakł. L. Fiszer, Warszawa–Łódź.
- Korczakowska M. 2001: *Polskie korpusy tekstów (wybrane zagadnienia)*, „Prace Językoznawcze” III, s. 65–75.
- Kosek I., Czerepowicka M., Przybyszewski S. 2020: „*Verbel*”. *Elektroniczny słownik paradygmatów polskich frazeologizmów czasownikowych. Teoria, problemy, prezentacja*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego w Olsztynie, Olsztyn.

41 <https://sgjw.clarin-pl.eu/about> (dostęp: 15 czerwca 2023).

- Kosek I., Przybyszewski S., Czerepowska M. 2018: *Źródła w tworzeniu elektronicznego słownika paradygmatów frazeologii-zmów werbalnych*, „Biuletyn Polskiego Towarzystwa Językoznawczego”, t. 74, s. 47–60.
- Ksw: *Kazania świętokrzyskie*, Rkpś 8001 BN, skany tekstu rękopiśmiennego dostępne na stronie biblioteki cyfrowej (<https://polona.pl/item/kazania-swietokrzyskie,MzAoOTIw/>), dostęp: 15 czerwca 2023).
- Kwaśnicka-Janowicz A., Pałka P. 2015: *Współczesny warsztat językoznawcy – elektroniczne zasoby i narzędzia badawcze*, „Język Polski” XCV, z. 1–2, s. 154–170.
- Kwaśnicka-Janowicz A., Pałka P. 2021: *Electronic language resource applications for Polish teachers in university teaching and teaching Polish as a foreign language* (współautor: Agata Kwaśnicka-Janowicz), [w:] S.J. Suárez Cuadros, E.J. Vercher García, B. Barros García, P. Marynenko, E.F. Quero Gervilla (red.), *New trends in Slavic studies – 2 / Современные исследования в славистике*: Выпуск 2 / Nuevas tendencias en estudios eslavos – 2, URSS, Moskva, s. 119–125.
- Language tools and resources for Polish* (<http://clip.ipipan.waw.pl/LRT>), dostęp: 7 czerwca 2023).
- NFJP: *Narodowy fotokorpus języka polskiego, 2014–2020* (<https://nfjp.pl>), dostęp: 15 czerwca 2023).
- NKJP: *Narodowy Korpus Języka Polskiego, 2008–2010* (<http://www.nkjp.pl/>), dostęp: 15 czerwca 2023).
- Nowak K. 2015: *Leksykografia elektroniczna. Praktyki, metody, narzędzia*, Instytut Języka Polskiego Polskiej Akademii Nauk (<https://scriptores.pl/wp-content/uploads/3.pdf>), dostęp: 6 czerwca 2023).
- Ogrodniczuk M., Nitoń B. 2020: *New developments in the Polish Parliamentary Corpus*, [w:] *Proceedings of ParlaCLARIN II Workshop*, Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020, s. 1–4 (<https://aclanthology.org/2020.parlaclarin-1.1.pdf>), dostęp: 20 października 2023).
- Ogrodniczuk M., Rudolf M., Wójtowicz B., Janicka S. 2022: *Error correction environment for the Polish Parliamentary Corpus*, [w:] *Proceedings of ParlaCLARIN III*, European Language Resources Association (ELRA), Marseille, 20 June 2022, s. 35–38 (<https://aclanthology.org/2022.parlaclarin-1.6.pdf>), dostęp: 20 października 2023).
- Open Multilingual Wordnet* (<https://omwn.org/>), dostęp: 7 czerwca 2023).
- Pałka P., Kwaśnicka-Janowicz A. 2017: *Przewodnik po elektronicznych zasobach językowych dla polonistów (słowniki, karteiki, korpusy, kompendia)*, Wydawnictwo JAK, Kraków.
- Pastuch M., Duda B., Lisczyk K., Mitrenga B., Przyklenk J., Sujkowska-Sobisz K. 2018: *Digital humanities in Poland from the perspective of the historical linguist of the Polish language: achievements, needs, demands*, „Digital Scholarship in the Humanities”, nr 33(4), s. 857–873.
- Pawłowski A. 2016: *Korpusy chronologiczne i leksykalne szeregi czasowe jako narzędzia wykrywania słów kluczowych i neosemantyzmów. Konceptualizacja czasu w korpusach chronologicznych*, „Przegląd Humanistyczny”, nr 60(3 (454)), s. 57–74.
- Pawłowski A. 2017: *ChronoPress – Chronological Corpus of Polish Press Texts (1945–1962)*, (https://www.clarin.eu/sites/default/files/Pawlowski-CLARIN2017_paper_35.pdf), dostęp: 20 października 2023).
- Piasecki M., Walkowiak T., Rudnicka E., Bond F. 2018: *Lexical platform – the first step towards user-centred integration of lexical resources*, „Cognitive Studies | Études cognitives”, nr 18, s. 1–20.
- Platforma leksykalna* (<http://lexp.clarin-pl.eu/>), dostęp: 15 czerwca 2023).
- Saloni Z., Gruszczyński W. 2013: *From multilingual to monolingual dictionaries. A historical overview of Polish lexicography*, „Studies of the Polish Linguistics”, t. 4, nr 8, s. 205–227.
- SAWA: *Słownik elektroniczny nazewnictwa Warszawy*, 2014 (<http://zil.ipipan.waw.pl/SAWA>), dostęp: 15 czerwca 2023).
- SEJF: *Słownik elektroniczny jednostek frazeologicznych*, 2015 (<http://zil.ipipan.waw.pl/SEJF>), dostęp: 15 czerwca 2023).
- SEJFEK: *Słownik elektroniczny jednostek frazeologicznych z ekonomii*, 2014 (<http://zil.ipipan.waw.pl/SEJFEK>), dostęp: 15 czerwca 2023).
- SFJP: W. Lubaszewski, H. Wróbel, M. Gajęcki, B. Moskal, A. Orzechowska, P. Pietras, P. Pisarek, T. Rokicka, *Słownik fleksyjny języka polskiego*, LexisNexis, Kraków 2001 (CD).
- SGJP 2012: Z. Saloni, M. Woliński, R. Wołosz, W. Gruszczyński, D. Skowrońska, *Słownik gramatyczny języka polskiego*, wersja 2.0, Sowa, Warszawa 2012 (CD).
- SGJP 2020: M. Woliński, Z. Saloni, R. Wołosz, W. Gruszczyński, D. Skowrońska, Z. Bronk, *Słownik gramatyczny języka polskiego*, wyd. 4 online, Warszawa 2020 (<http://sgjp.pl/leksemy/>), dostęp: 15 czerwca 2023).
- SGPKa: J. Karłowicz, *Słownik gwar polskich*, t. 1–6, Akademia Umiejętności, Kraków 1900–1911, skany tekstu drukowanego dostępne na stronie biblioteki cyfrowej (<https://zbc.uz.zgora.pl/dlibra/publication/9017>), dostęp: 15 czerwca 2023).
- SJP PWN 2023: *Słownik języka polskiego PWN* (<http://sjp.pwn.pl/sjp/lista/A.html>), dostęp: 15 czerwca 2023).

- Skowronek K., Czopek-Kopciuch B., Górny H., Magda-Czekaj M., Palinciuc-Duek E., Supranowicz E. 2018: *Najczęstsze nazwiska w Polsce – współczesność i historia. Słownik elektroniczny. Stan badań nad projektem*, „Onomastica” LXII, s. 299–314.
- Słownik synonimów i antonimów*, red. P. Żmigrodzki, Europa, Wrocław 2005 (<http://leksykony.interia.pl/synonim>, brak dostępu).
- Słownik wyrazów obcych i zwrotów obcojęzycznych Władysława Kopalińskiego* (<http://www.sloownik-online.pl/>, brak dostępu).
- Słowniki dawne i współczesne* (<http://www.leksykografia.uw.edu.pl/>, dostęp: 15 czerwca 2023).
- Słowosieć: Wielki relacyjny słownik semantyczny języka polskiego*, wersja 4.2., 2023 (<http://plwordnet.pwr.wroc.pl/wordnet/>, dostęp: 15 czerwca 2023).
- SNP: *Słownik neologizmów polskich*, 2014 – praca w toku (<https://obserwatoriumjezykowe.uw.edu.pl/czym-jest-sloownik-neologizmow-polskich/>, dostęp: 15 czerwca 2023).
- SPXVI: *Słownik polszczyzny XVI wieku*, red. M.R. Mayenowa, F. Peplowski, K. Mrowcewicz, P. Potoniec, t. 1–39 (praca w toku), Zakład Narodowy im. Ossolińskich, Wrocław–Warszawa 1956–2021 (<https://spxvi.edu.pl/>, dostęp: 15 czerwca 2023).
- ST WSJP PAN: *Słownik tematyczny WSJP PAN*, 2021 – praca w toku (https://wsjp.pl/sloownik_tematyczny, dostęp: 15 czerwca 2023).
- SWJP: *Słownik współczesnego języka polskiego*, red. B. Dunaj, Wilga, Warszawa 1996; *Słownik współczesnego języka polskiego*, red. B. Dunaj, t. 1–2, Wilga, Warszawa 1999; *Współczesny słownik języka polskiego*, red. B. Dunaj, t. 1–2, Langenscheidt Polska, Warszawa 2007 (tytuł słownika zmieniono bez wiedzy i zgody redaktora oraz autorów).
- SYNAMET: *Mikrokorpus metafor synestezyjnych*, 2019 (<http://synamet.polon.uw.edu.pl/korpus>, dostęp: 15 czerwca 2023).
- USJP 2003: *Uniwersalny słownik języka polskiego*, red. S. Dubisz, t. 1–4, Wydawnictwo Naukowe PWN, Warszawa 2003.
- USJP 2004: *Uniwersalny słownik języka polskiego*, red. S. Dubisz, t. 1–4, Wydawnictwo Naukowe PWN, Warszawa 2003; wersja 1.0, Wydawnictwo Naukowe PWN, Warszawa 2004 (CD).
- USJP 2016: *Uniwersalny słownik języka polskiego* (<https://usjp.pwn.pl/>).
- Verbel: Słownik paradygmatów polskich frazeologizmów czasownikowych*, oprac. I. Kosek, M. Czerepowicka, S. Przyby-szewski, 2018 (<http://uwm.edu.pl/verbel/>, dostęp: 15 czerwca 2023).
- Walenty: Słownik walencyjny języka polskiego Walenty*, 2021 (<http://walenty.ipipan.waw.pl/>, dostęp: 15 czerwca 2023).
- Wawrzyńczyk J., Wierzchoń P., Borchmann Ł. 2021: *Sześćdziesiąt lat Narodowego Fotokorpusu Języka Polskiego*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.
- Werla M., Maryl M. (oprac.) 2014: *Humanistyczne projekty cyfrowe w Polsce. Wersja 1.1* (26 czerwca 2014), Poznań–Warszawa (https://rcin.org.pl/Content/65767/PDF/WA248_84544_bez-sygn_werla-maryl-human_o.pdf, dostęp: 15 czerwca 2023).
- Wierzchoń P., Nowak P. 2020: *Digital libraries and the breakthrough in linguistic chronologization. The applications of digitization in linguistics*, „Zagadnienia Informatyki Naukowej. Studia Informacyjne”, nr 58(1A), s. 110–121.
- Wiktoria: *Słownik onomazjologiczny*, 2009–2011 (<https://wiktoria.ijp.pan.pl/>, dostęp: 15 czerwca 2023).
- WSJP PAN: *Wielki słownik języka polskiego PAN*, red. P. Żmigrodzki, 2007 – praca w toku (<https://wsjp.pl/>, dostęp: 15 czerwca 2023).
- WSJP PWN 2018a: *Wielki słownik języka polskiego PWN ze słownikiem wyrazów bliskoznacznych*, Wydawnictwo Naukowe PWN, 2018 (pendrive).
- WSJP PWN 2018b: *Wielki słownik języka polskiego PWN*, red. S. Dubisz, t. 1–5, Wydawnictwo Naukowe PWN, Warszawa 2018.
- Wyszukiwarka Poliqarp dla NKJP: *Narodowy Korpus Języka Polskiego*, 2008–2010 (<http://nkjp.pl/poliqarp/>, dostęp: 15 czerwca 2023).
- Zawisławska M. 2019: *Metaphor and senses: The Synamet Corpus. A Polish resource for synesthetic metaphors*, Peter Lang, Berlin–Bern–Bruxelles–New York–Oxford–Warszawa–Wien.
- Żmigrodzki P. 2022a: *Wielki słownik języka polskiego PAN jako słownik tematyczny*, „Studia Językoznawcze”, t. 21, s. 153–166.
- Żmigrodzki P. 2022b: *Wielki słownik języka polskiego PAN w 2022 roku – osiągnięcia, perspektywy, wyzwania*, „Język Polski” CII, z. 3, s. 129–136.

Summary

New electronic language resources for Polish philologists (dictionaries and corpora)

Keywords: digital humanities, dictionary, text corpus, electronic language resources.

Polish-language resources of digital humanities have been growing and developing very rapidly over the last several years. On the one hand, such a large and constantly growing number of various electronic language resources has a significant impact on the development of scientific research and didactic methods. On the other hand, this situation makes it difficult for researchers to be up-to-date on news and changes that are taking place in already existing digital tools, dictionaries, corpora, libraries or compendia. Therefore, it is necessary to update the studies on a regular basis in which various electronic language resources are discussed in detail. One of the studies that needs to be supplemented and corrected is *Przewodnik po elektronicznych zasobach językowych dla polonistów* (*Guide to electronic language resources for polonists (dictionaries, card catalogs, corpora, compendia)*) by P. Pałka and A. Kwaśnicka-Janowicz (2017). For that reason, in this paper we will focus on selected electronic language resources that have not been included in the mentioned book and which refer to linguistic data dating from 1945 to the present or covering the entire 20th century. In the summary, we will discuss possible ways of using electronic language resources (selected for the paper) in research and teaching.