

JOLANTA KLIMEK-GRĄDZKA

KORPUSY JĘZYKOWE A/I SEMANTYKA HISTORYCZNA – *FORMAT*

Wojciech Chlebda stwierdził, że „[d]ociekanie zawartości pojęć (czy, szerzej, konceptów) powinno być poprzedzone utworzeniem materiałowego fundamentu, na którym wznoszona będzie architektura eksplikacji” (Chlebda, 2016, s. 328), i tę myśl chcę uczynić fundamentem rozważań semantycznych, łączących badania historycznojęzykowe i korpusowe polszczyzny. Uznając te dwie metody badań lingwistycznych za dopełniające się, wychodzę od analiz właściwych semantyce historycznej (Sieradzka-Baziur, 2017) dla ustalenia przebiegu zmian znaczeniowych leksemu, po czym chcę je skonfrontować z materiałem pozyskanym z korpusów językowych. Taki sposób oglądu materiału pozwoli sprawdzić, w jakim stopniu dane z różnych typologicznie źródeł są zbieżne, w jakim zaś dane korpusowe wyprzedzają słownikowe i dzięki temu mogą być wykorzystane do śledzenia neosemantyzmów *in statu nascendi*.

Badania oparte na ekscerpcji słownikowej (SPXVI, SWil, SL, SW, SJPD, USJP, WSJP PAN) i korpusowej (Odkrywka, Chronopress, SpokesPL, MoncoPL, Narodowy Korpus Języka Polskiego i Korpus Dyskursu Parlamentarnego) oraz materiałach własnych autorki (prasa, tv, Internet) sprowokowały szerzące się, zwłaszcza w odmianie publicystycznej i politycznej polszczyzny współczesnej, połączenia typu *format normandzki*, *format wyszehradzki*. Celem badań jest jednak nie tylko obserwacja wykształcania się nowego znaczenia leksemu *format*, ustalenie jego synonimów tekstowych i chronologizacja, ale przede wszystkim rozpoznanie możliwości odkrywania nowych znaczeń wyrazu przy wykorzystaniu danych i narzędzi, oferowanych przez humanistykę cyfrową (*computing in the humanities*; por. Bomba, 2013).

Dr hab. JOLANTA KLIMEK-GRĄDZKA, prof. KUL – Katolicki Uniwersytet Lubelski Jana Pawła II, Wydział Nauk Humanistycznych, Instytut Językoznawstwa; adres do korespondencji: Al. Raławickie 14, 20-950 Lublin; e-mail: jolanta.klimek-gradzka@kul.pl; ORCID: <https://orcid.org/0000-0002-6289-5634>.

Artykuły są objęte licencją Creative Commons Uznanie autorstwa – Użycie niekomercyjne – Bez utworów zależnych 4.0 Międzynarodowe (CC BY-NC-ND 4.0)

FORMAT – HISTORIA SŁOWA

Pierwsze notacje interesującego mnie rzeczownika pochodzą z drugiej połowy XVI wieku i co ciekawe, dotyczą formy pluralnej. SPXVI za słownikiem Knapiusza notuje *formaty* z Piotra Skargi z objaśnieniem: ‘list polecający, zaświadczenie biskupie, że kapłan udający się poza diecezję jest wolny od cenzur kościelnych i może odprawiać mszę św.’:

Dziś siedzi ná niey [na stolicy rzymskiej] Syricius/ ktory nam towárzyfzem ieft/ fktorym známi świat wżyftek przez połpolitowanie formátow w iednego towárzyftwá fpołeczności zgadza fię.

W *Elektronicznym słowniku XVII i XVIII w.* interesującego mnie hasła brak, jednak cytaty w Elektronicznym korpusie tekstów polskich z XVII i XVIII wieku (KorBa) dowodzą istnienia znaczenia zbieżnego z tym sprzed stulecia:

Daie Krol formaty (niby to Prezentę) temi słowy do Biskupa. Abyś w twey Diecezyi dawał swięcenie, daiemy ci władzę tym listem (B. Chmielowski);

Regestr pretensyj powinien był być spisany w formacie rejestrowej, podpisany ręką podającego bądź pana, bądź plenipotentą i pieczęcią wielką herbową pańską podpieczętowany (J. Kitowicz),

choć w tym ostatnim wypadku widać już, że *format* – czy *ta formata* – przybiera formę liczby pojedynczej.

Interesujące jest, że Linde, tworzący w początkach XIX wieku swój słownik, nie zanotował *formatu* jako osobnego hasła – czy można to odczytywać jako dowód na znikomą frekwencję słowa, pozostawiam bez rozstrzygnięcia. Niewątpliwie polszczyzna słowo znała, co więcej – w odmiennym od dotychczas notowanego znaczeniu, skoro Linde zamieścił *format* w artykule objaśniającym rzeczownik *forma* jako synonim *formy książki* (SL 1, s. 652). Analiza źródeł wskazuje, że już w latach czterdziestych XVIII wieku *format* w znaczeniu ‘rozmiar książki’ funkcjonował w obiegu, a nawet wprowadzany był do tytułów publikacji, np.:

Wybrany y Zupełny Kancyonał Gdanski, w Którym 565. Pieśni Duchownych [...] Z przydatkiem Pieśni y nabożnych Modlitew Jako też y krotkich Pytań Katechizmowych. Na wielu zadanie Pierwszym razem w tym wygodnym formacie wygotowany.

To nowe znaczenie zaczęło się szybko upowszechniać, tak że już w słowniku tzw. wileńskim z 1861 roku *format* to przede wszystkim ‘forma, kształt książki’, a dopiero wtórnie w znanym od wieków znaczeniu i formie pluralnej *formaty*

‘list świadeczny od biskupa do innej djecezji’ (SWil). Identycznie rzecz wygląda w tzw. słowniku warszawskim, gdzie *format* to ‘forma, wielkość, wymiary książki’, obok tego *formaty* blp. ‘świadectwo od biskupa do innej djecezji, świadectwo święceń’. Drugie znaczenie zostało już pominięte w *Słowniku języka polskiego* pod redakcją Witolda Doroszewskiego – tu *format* tylko w sensie ‘rozmiar, wielkość, kształt, wymiar (np. książki, arkusza)’ (SJPD). Sygnałem wykształcania się dodatkowego znaczenia przenośnego jest notatka umieszczona w artykule hasłowym: „ludzie też mają format”. Potencjał specjalizacyjny tkwiący w sformułowanej przez Doroszewskiego definicji leksemu, wskazującej na różne typy opisu fizycznego przedmiotu, a także nieograniczającej desygnatu opisywanego za pomocą *formatu*, został wykorzystany w kolejnych słownikach.

Leksykon pod redakcją Mirosława Bańki wydziela wprawdzie wciąż dwa znaczenia: dosłowne 1) format książki, zdjęcia, papieru itp. to ich rozmiary, przez drukarzy i wydawców określane za pomocą specjalnych symboli, i przenośne (choć bez takiej kwalifikacji eksplicytnie wyrażonej), 2) ktoś lub coś jakiegoś formatu, np. wielkiego lub małego, to osoba lub rzecz oceniana w taki sposób (ISJP 1, s. 416), lecz bliski mu czasowo *Słownik współczesnego języka polskiego* pod redakcją Bogusława Dunaja wskazuje dwa dosłowne sensy, wyrastające z dotychczasowego jednego: 1) ‘rozmiar, wielkość czegoś, szczególnie książek, zeszytów, arkuszy papieru’, 2) *inform.* ‘ściśle ustalony sposób zapisu określonego typu informacji na dysku komputera’, poza tym znany już {ktoś} {jakiegoś} *formatu* ‘o człowieku: reprezentujący wysoki poziom uzdolnienia, charakteru itp.’ (SWJP 1, s. 256). *Praktyczny słownik współczesnej polszczyzny* pod redakcją Haliny Zgółkowej wprawdzie wyróżnia dwa znaczenia, jak dotychczas wymienione leksykony, ale nieco inaczej je formułuje: 1) ‘wymiar i kształt określonej powierzchni’, 2) przenośny ‘miara, poziom’ (PSWP 11, ss. 407-408). Wciąż powstający i nieustannie aktualizowany na podstawie danych korpusowych *Wielki słownik języka polskiego* pod redakcją Piotra Źmigrodzkiego odnotowuje już sześć znaczeń *formatu*, wyraźnie wyspecjalizowanych, na co jednoznacznie wskazują towarzyszące definicjom kwalifikatory:

- 1) ‘rozmiar arkusza papieru lub czegoś, co na nim zostało wydrukowane lub umieszczone w inny sposób’
- 2) książk. ‘poziom, stopień doskonałości, wybitności kogoś lub czegoś’
- 3) *inform.* ściśle określony sposób zapisu danych przez systemy komputerowe
- 4) *inform. pot.* formatowanie dysku w komputerze
- 5) tel. licencja na wyprodukowanie i emisję w telewizji danego kraju programu lub serialu, który jest lub był produkowany i nadawany w telewizji w innym kraju

6) rad. taka konstrukcja programu radiowego, przede wszystkim dobór gatunków odtwarzanej muzyki i tematów poruszanych na antenie, aby był on atrakcyjny dla grupy słuchaczy, jakich stacja chce pozyskać ze względu na cele komercyjne i reklamowe.

Nagły przyrost znaczeń wyspecjalizowanych do poszczególnych działów mediów i techniki wiązać można z rozwojem technologicznym (co skutkuje możliwością zapisu danych nie tylko na nośnikach „papierowych”) oraz globalizacją produkcji telewizyjnych. Mimo dostrzegalnego rozdrobnienia znaczeń leksemu *format* w *Wielkim słowniku języka polskiego* trudno odnaleźć wśród nich sens aktualizowany w dających się słyszeć w mediach i stanowiących, jak wspomniano we wstępie, asumpt do podjęcia niniejszych badań, połączeniach typu *format normandzki*.

FORMAT – FREKWENCYJNE DANE KORPUSOWE

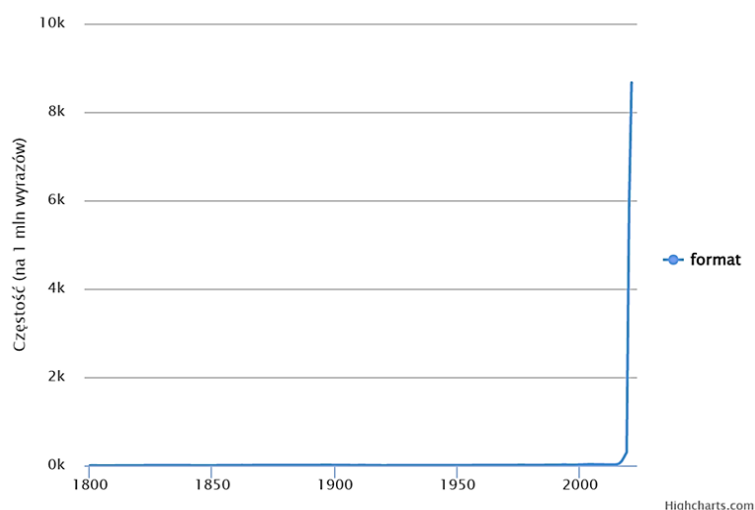
Jak piszą R.L. Górski, M. Król i M. Eder:

[j]ęzykoznawca badający przeszłość języka otrzymuje coraz to nowe narzędzia – dotąd dysponował naukową edycją tekstu oraz słownikiem wraz ze stworzoną do niego kartoteką. I choć edycja i słownik nadal zajmują poczesne miejsce w instrumentarium lingwisty, przestały być jedynymi dostępnymi badaczowi narzędziami. (Górski i in., 2019, s. 7)

W poszukiwaniu danych uzupełniających dane słownikowe sięgnęłam więc do korpusów, zawierających słownictwo z różnego czasu i różnych odmian stylowych języka. Z korpusów: Odkrywka, Chronopress, SpokesPL, MoncoPL, Narodowy Korpus Języka Polskiego i Korpus Dyskursu Parlamentarnego wydobylam nie tylko użycia leksemu *format* w połączeniach typu *format wyszehradzki*, *format normandzki*, ale także dane ilościowe, pokazujące częstotliwość występowania tychże.

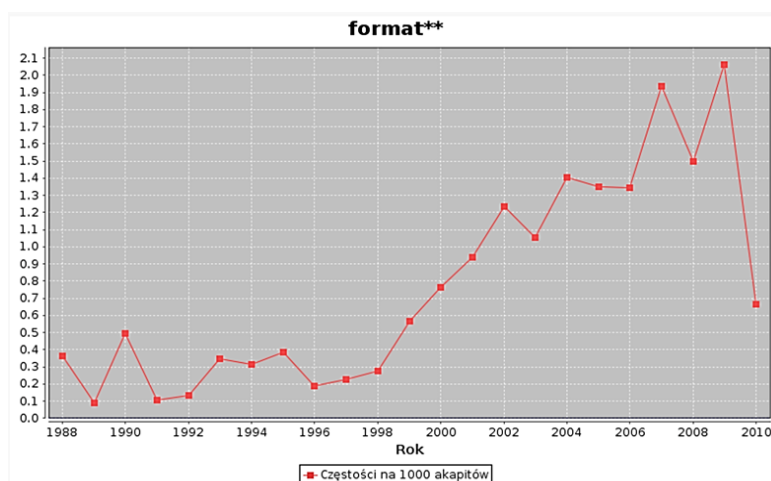
Dane pozyskane z Odkrywki¹, korpusu (s)tworzonego przez Filipa Gralińskiego (2019) i gromadzącego dane językowe od 1800 roku, jednoznacznie wskazują narastającą, w ostatnich latach wręcz lawinowo, popularność leksemu *format*:

¹ Za możliwość skorzystania z Odkrywki składam serdeczne podziękowania jej Twórcy.

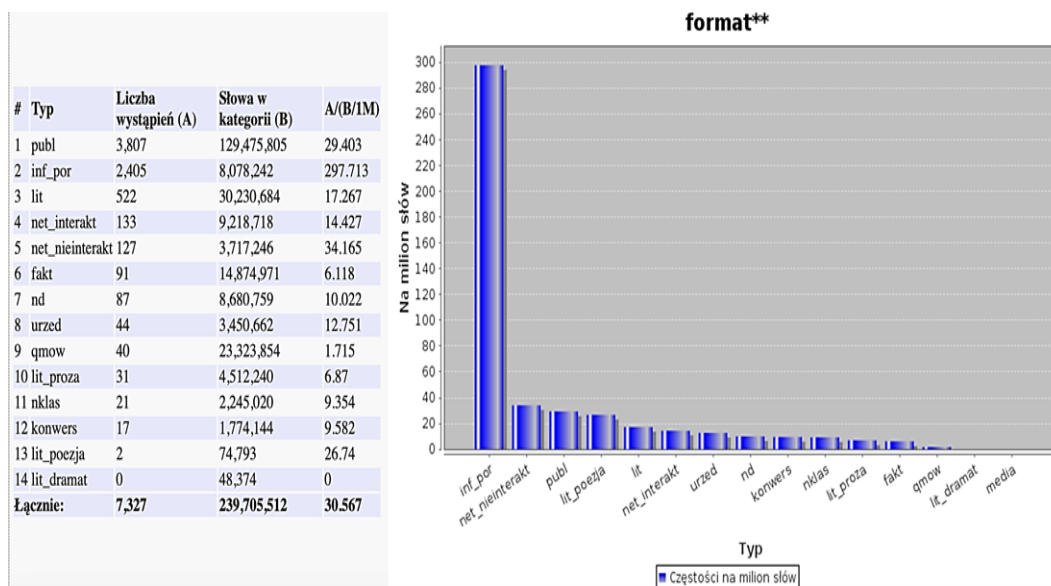
Wykres 1. *Format* (źródło: Odkrywka)

Skokowy przyrost w ostatnich dekadach – z kilkunastu-kilkudziesięciu do kilku tysięcy użyć na milion słów – nie jest zaskoczeniem, gdy sięgnąć do przytoczonych wcześniej danych leksykograficznych i odnotowanych w nich wyodrębnionych znaczeń medialnych i informatycznych. Zwłaszcza te ostatnie, wydostawszy się z języka specjalistycznego, musiały wpłynąć na częstotliwość użycia słowa.

Stale przyrastanie frekwencji *formatu* poświadcza też wyszukiwarka PELCRA Narodowego Korpusu Języka Polskiego:

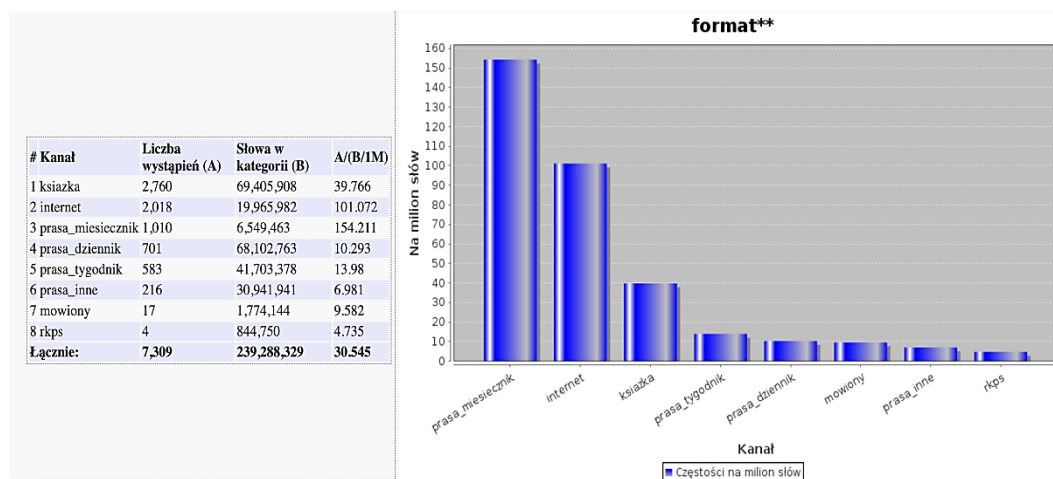
Wykres 2. *Format* i jego formy fleksyjne w latach 1988-2010 (źródło: NKJP)

NKJP dostarcza też informacji, że zasięg stylowy *formatu* ogranicza się *de facto* do tekstów publicystycznych (52%), informacyjno-poradnikowych (33%), którym znacznie ustępuje literatura (7%).

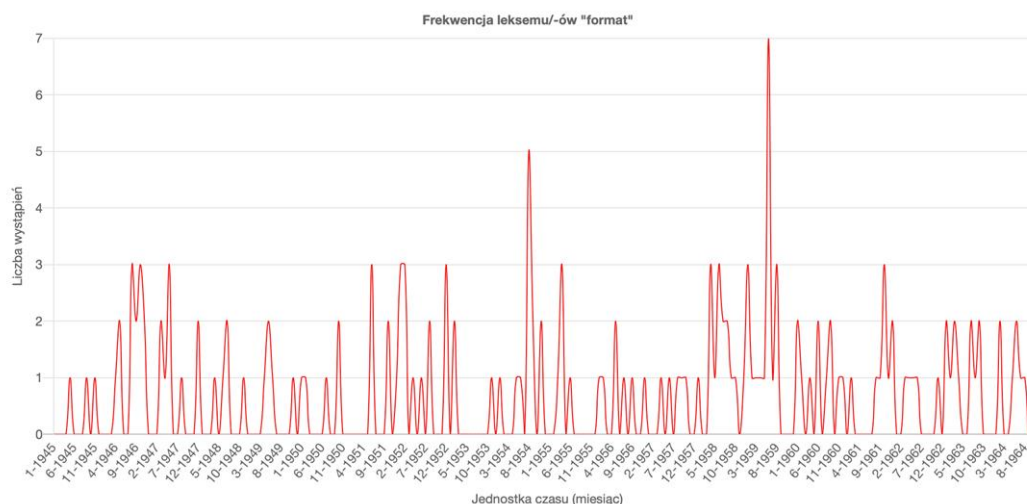


Wykres 3. *Format* – częstość w zależności od stylu źródła (źródło: NKJP)

Gdy spojrzeć na zależność pomiędzy częstością użycia *formatu* a typem tekstu (#kanał w taksonomii NKJP; Przepiórkowski i in., 2012), okazuje się, że najczęściej ze słowem można się spotkać w książkach (38%), prasie (sumarycznie w dziennikach, miesięcznikach i innych periodykach 34%) i Internecie (28%). Zaskakująca jest znikoma obecność *formatu* w potocznym języku mówionym, intuicja podpowiadała bowiem, że podstawowe znaczenia są wykorzystywane także w tej odmianie polszczyzny.

Wykres 4. *Format* – częstość w zależności od typu źródła (źródło: NKJP)

Dane pozyskane z NKJP zestawione z danymi z korpusu ChronoPress (Pawłowski, 2021) pokazują, że w ciągu minionych dekad zmieniła się częstotliwość użycia w prasie słowa *format*, co bezpośrednio wynika ze zmiany znaczenia i gwałtownego przyrostu sensów specjalistycznych powiązanych z rozwojem mediów i nowych technologii.

Wykres 5. *Format* w prasie 1945-1964 (źródło: ChronoPress)

W prasie z lat 1945-1964 *format* pojawił się zaledwie kilkadziesiąt razy, głównie w kontekście książek, artystów, okien czy telewizora, co odpowiada ówczesnym znaczeniom leksemu. Kilka przykładów z ChronoPressu:

Tymczasem wszystko, co da się pokazać w cinemascope, da się pokazać także na ekranie dawnego formatu, zwanym obecnie za granicą pogardliwie ekranem formatu naparstka.

Wreszcie w finale spotyka się z pięściarzem wielkiego formatu, Murzynem Sandersem.

Wanny, umywalnie, płuczki, rury, piecyki, wszelkiego rodzaju i wszelkiego formatu kształtki do instalacji wodnych – oto materiały niezbędne dla budownictwa, które odnajdujemy przygotowane do ekspedycji w magazynach Centrali Handlowej Przemysłu Metalowego w Warszawie.

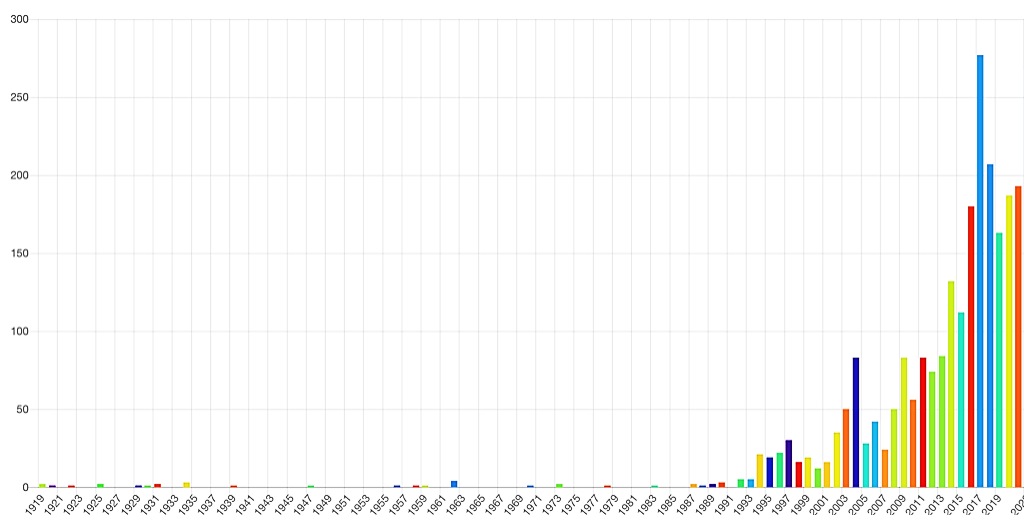
Jak powiedziano wcześniej, współczesne słowniki, zwłaszcza WSJP, wydzielają specjalistyczne znaczenia *formatu* skupione wokół szeroko rozumianych technologii (informatycznych czy medialno-produkcyjnych). To po części może wyjaśniać rozkład wyników pozyskanych z korpusu polszczyzny mówionej SpokesPL (Pęzik, 2015). *Format* pojawia się wyłącznie w wypowiedziach ludzi w wieku 20-60 lat – przy czym najczęściej z przedziału 40-50 lat, z wyższym wykształceniem, w wypowiedziach mężczyzn aż trzykrotnie częściej niż kobiet, np.:

jeden giga ale to jeden giga to w tym formacie to ma pięćset zdjęć się zmieścić,

w ogóle system stwierdza że to jest bardzo dziwny format pliku,

o zależy w jakiej jako jakości w jakim formacie zapisujesz w JPG na pewno zmieści ci się mało.

Ostatnim źródłem, z którego zaczerpnęłam dane frekwencyjne, jest Korpus Dyskursu Parlamentarnego (Ogrodniczuk, 2018). Tu wyniki przedstawiają się zgoła odmiennie, niż w omówionych dotychczas korpusach. Jak pokazuje wykres 6, w języku szeroko rozumianej polityki *format* jest słowem dość często notowanym:

Wykres 6. *Format* w Korpusie Dyskursu Parlamentarnego (źródło: KDP)

Poświadczenia użycia leksemu w dyskursie parlamentarnym istnieją od roku 1919², choć zauważalny jest ich regularny przyrost od połowy lat dziewięćdziesiątych, gwałtowny zwłaszcza w ostatniej dekadzie³. Szczególnie interesująco przedstawiają się lata 2017 i 2022, w których leksem i jego formy fleksyjne zanotowano odpowiednio: 277 i 267 razy, co stanowi zauważalne odstępstwo od parlamentarnej przeciętnej. Zastrzeżenie do metadanych pozwala wyjaśnić tę nagłą popularność *formatu*: w 2017 roku w Komisji Kultury i Środków Przekazu dyskutowano nad formatami medialnymi itp., zaś w 2022 roku nad kwestiami polityki zagranicznej. W pierwszym wypadku dominują więc połączenia typu: *format telewizyjny*, *format radiowy*, *format festiwalu*, odsyłające do wyodrębnionych w WSJP znaczeń leksemu centralnego, w drugim: *format Trójkąta weimarskiego*, *format weimarski*, *formaty regionalne*, wyzyskujące nienotowany w słownikach sens rzeczownika *format*.

FORMAT – KOŁOKACJE

Dane korpusowe pokazują, że istotnie pojawia się w polszczyźnie pomijane przez słowniki znaczenie słowa *format*. Warto to wstępne rozpoznanie potwierdzić, jak

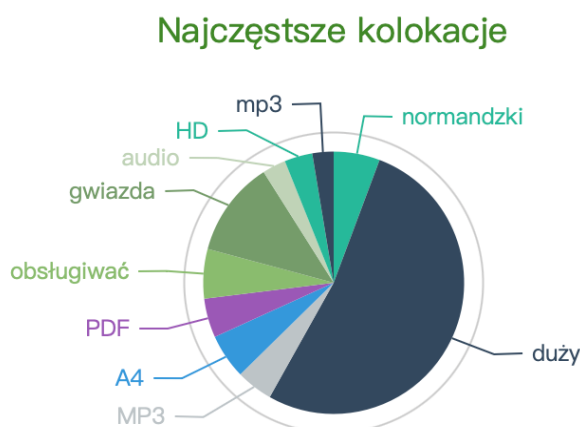
² Stan na dzień 7 lipca 2023 roku.

³ Trzeba pamiętać, że KDP, w przeciwieństwie do NKJP, nie jest korpusem zrównoważonym, lecz gromadzi wszystkie wystąpienia sejmowe od roku 1919.

bowiem uważa Renata Przybylska, „badanie kolokacji przesądza też o wykrywaniu nowych znaczeń wyrazu [...] w tym zwłaszcza tych mających motywację metaforyczną” (Przybylska, 2020, s. 49). Taki sposób dociekania sensu zbliża nas do badań z zakresu semantyki historycznej, które wobec braku możliwości

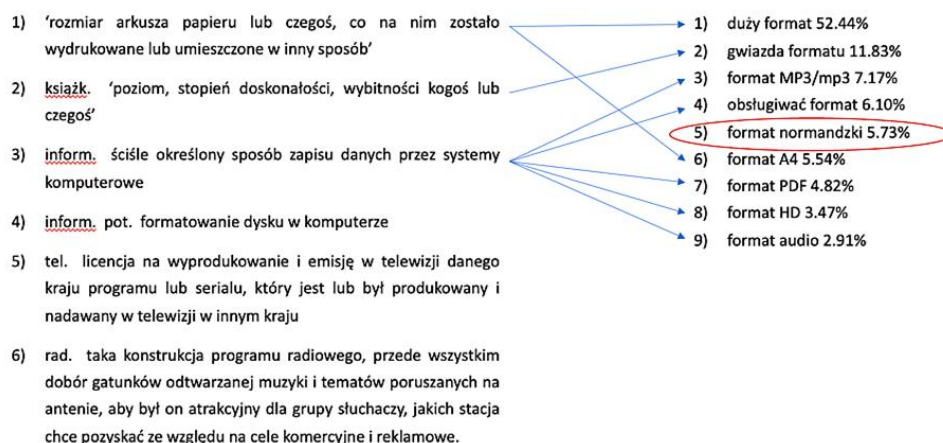
odwołania się do kompetencji językowej badacza jako instancji weryfikującej, wymagają twardego oparcia w postaci danych tekstowych. [...] Tylko w ten sposób, a więc wyzyskując informacje kontekstowe do ustalania zawartości tekstowej słów, można wyjść poza sferę domysłów i przypuszczeń. (Woźniak, 2011, ss. 71-72)

Kolokacje, rozumiane za Przybylską jako wszelkie połączenia tekstowe wyrazu z innymi wyrazami (Przybylska, 2007, s. 69), leksemu *format* wyekstrahowane z korpusu monitorującego MoncoPL (Pęzik, 2020) przedstawiają się następująco:



Wykres 7. Kolokacje leksemu *format* (źródło: MoncoPL)

Aż 52.44% wszystkich notacji przypada na *duży format*, na drugim miejscu ze znacznie niższą częstością pojawia się *gwiazda (jakiegoś) formatu* (11.83%), na trzecim *format MP3* (traktowany łącznie z *mp3*: 7.17%), dalej kolejno: *obsługiwać format* (6.10%), *format normandzki* (5.73%), *format A4* (5.54%), *format PDF* (4.82%), *format HD* (3.47%) i *format audio* (2.91). To dane wyjątkowo cenne, bowiem, jak piszą twórcy MoncoPL, „mogą być przydatne w identyfikacji dominujących lub wręcz powstających w danym czasie sensów wyrazów wieloznacznych” (Pęzik, 2020, s. 145). Mając do dyspozycji stale rozwijany WSJP i stale aktualizowany korpus MoncoPL, warto moim zdaniem zestawić definicje słownikowe z danymi uzualnymi:



Wykres 8. Relacja pomiędzy danymi słownikowymi (WSJP) a korpusowymi (MoncoPL); oprac. własne

Trzy pierwsze wymienione przez słownik znaczenia okazują się najczęściej wykorzystywane w języku – to nie tylko nie dziwi, ale i potwierdza słusność zaproponowanego przez redaktorów WSJP uszeregowania sensów w słowniku. Uwagę natomiast zwraca grupa imienna, której członu głównego nie sposób przyporządkować do żadnego z wyodrębnionych sensów: *format normandzki*. Zgodnie z przyjętą za Przybylską metodą weryfikowania istnienia/powstania nowego sensu przenośnego wyrazu, warto bliżej przyjrzeć się temu połączeniu, które uznać należy za „poznawczo istotne dla semantyki badanego wyrazu hasłowego” (Przybylska, 2020, s. 46).

Przeszukanie konkordancji *formatu* w NKJP i KDP przynosi co najmniej zaskakujące wyniki. *Format* pojawia się bardzo często w kolokacjach tożsamych lub zbliżonych do wspomnianego już *formatu normandzkiego*. W wymienionych korpusach dają się zaobserwować zestawienia *formatu* z:

- 1) przymiotnikami, np.: *wielostronny, bałtycki, wyszehradzki, unijny, miński, dwustronny, weimarski, transgraniczny, negocjacyjny*;
- 2) rzeczownikami, np.: *szeferów państw, Trójmorza, Eurogrupy, telefonicznej konferencji, dyskusji, rozmów, współpracy, do spraw praw człowieka, podstolika „Nauczyciel w systemie edukacji”, małej kwadrygi, Grupa Wyszehradzka Plus, Ligi Narodów*; te połączenia mogą przybierać także postać rozbudowanych szeregów i fraz, np.: [w formacie] *Unia Europejska, Polska – z jednej strony, i Rosja – z drugiej*; [w formacie] *z jednej strony – Komisja, Polska, z drugiej strony – Federacja Rosyjska; polsko-niemieckiego okrągłego stołu*.

Wykorzystanie opisanego przez Przybylską (2020, s. 47) sposobu identyfikowania nowych znaczeń rzeczowników drogą obserwacji ich kolokacji prowadzi do wniosku, że istnieje dodatkowy sens leksemu *format*. Notowane w korpusach określenia przyłączone przez *format* wskazywać mogą na cel działania *formatu*, czynności wykonywane przez *format*, liczbę lub liczbę stron uczestniczących w *formacie*, a także precyzować te strony. To ostatnie dokonuje się albo poprzez eksplcytne wyliczenie stron *formatu*, albo odwołanie do nazwy gremium formalnie istniejącego w relacjach międzynarodowych. Taki zakres występowania nowego *formatu* mógłby dostarczyć cennych wskazówek co do jego genezy na gruncie polskim, pozwalając go wywieść od angielskiego *Normandy Format* (to jednocześnie jedyna pasująca do wzorca kolokacja notowana w NOW Corpus), gdyby nie to, że *format* w przenośnym znaczeniu pojawił się w języku polskim wcześniej, niż została powołana w 2014 roku dla rozstrzygnięcia kwestii dotyczących wojny w Donbasie i przynależności Krymu grupa złożona z przedstawicieli Ukrainy, Francji, Niemiec i Rosji.

KDP w latach 1926-2014 odnotowuje *format* prawie tysiąc razy: w odniesieniu do człowieka (*ludzie wielkiego formatu*, *wielkiego formatu intelektualnego*, *politycy wielkiego formatu*, *coś świadczy o formacie szefa rządu*, *uczni światowego formatu*, *ludzie właściwego*, *dużego formatu*, *to nie ten format dostojnika*, *wielkiego formatu pianista*, *osoby formatu np. ojca Rydzyka*, *wybitny mąż stanu o formacie międzynarodowym*), rozmiaru (np. tomów, książek, prasy, płacht, banknotów, sprawozdań, obrazu, fiszek, ustaw, ale i zjawisk – *patologia społeczna wielkiego formatu*, *klęski żywiołowe najróżniejszego formatu*, *rozwiązania światowego formatu*, *unijny*, *jednolity format wizowy*, *konstytucyjny format rodziny*) czy sposobu zapisu (*format pliku*, *format bazodanowy*, *format tekstowy*, *format MP3*, *format SD*). Tylko w pojedynczych notacjach pojawia się *format* w sensie mnie interesującym, np.: *ustrój o zdecydowanie centralistycznym formacie* (1996), *nastąpi zatem powrót do formatu, do kształtu dyrektywy przyjętej wcześniej przez Parlament Europejski* (2004), *W dniu 20 lipca w Warszawie odbyły się rozmowy z przedstawicielami DG SANCO i DG Trade, przygotowujące rozmowy w formacie Unia Europejska, Polska – z jednej strony, i Rosja – z drugiej, mające na celu zniesienie wprowadzonych przez Rosję zakazów importu z Polski produktów pochodzenia roślinnego i zwierzęcego* (2006), *Nie było takich, biorąc na poważnie, kontaktów, jak się to mówi w języku dyplomacji, w tym formacie między na przykład premierem Wielkiej Brytanii czy prezydentem Francji i szefem rządu w Polsce* (2007), *Bez takiej rozmowy będzie bardzo trudno zakroić nową formułę, nowy format współpracy bez przeszkód i format realizowania nowej wizji Unii Europejskiej* (2007), *powinno już dojść do wypełnienia szeregu*

zalożonych celów w stosunkach dwustronnych oraz do przeprowadzenia w formacie wielostronnym odpowiednich działań finansowych (2009), Grupa Wyszehradzka przyciąga dzięki swojej efektywności kolejne państwa – dzięki inicjatywie pana ministra Sikorskiego coraz częściej funkcjonuje w formacie Grupa Wyszehradzka Plus, przyciągając kolejne państwa regionu (2012). Przykłady te wskazują, że poszerzanie się abstrakcyjnego sensu *formatu* nie jest bezpośrednio związane z wydarzeniami roku 2014 i powołaniem Formatu Normandzkiego. Niewątpliwie jednak rok ten stanowi istotną cezurę w dziejach leksemu, czego dowodem jest fakt, iż KDP w latach 2014-2021 notuje aż 1698 wyników (stan na 5 grudnia 2022), z czego duża część przypada na *format* w znaczeniu odnoszącym się do telewizji (*musimy mieć po prostu polskie formaty, które będą opowiadać o polskiej historii. I czymś takim będzie „Korona królów”*, 2017; *No i kupiliśmy ten format [o teleturnieju „Jaka to melodia?”]*, 2017), obok tego jednak równoprawnie niezdefiniowany nowy *format*, np.:

Podczas ostatniego posiedzenia w **formacie do spraw praw człowieka**, które odbyło się w ubiegłym tygodniu, Komitet Ministrów zdecydował też o zakończeniu nadzoru nad wykonywaniem wyroków.

Format „16+1” nie jest w żaden sposób ani zagrożeniem dla jedności unijnej, ani podważeniem zasad unijnych. Staramy się to naszym kolegom z Niemiec i z innych krajów wyjaśniać. I mam wrażenie, że to zrozumienie coraz bardziej się tam pojawia. To jest chyba po prostu nieporozumienie związane z relatywnie nowym charakterem tego **formatu**, niemniej jednak jest to **format** przejrzysty. My zapraszamy przedstawicieli Komisji Europejskiej do uczestnictwa w tym **formacie**, a ci przedstawiciele Komisji Europejskiej robią notatkę, która jest dostępna dla państw członkowskich. Uważamy, że robimy wszystko, żeby pokazać, że współpraca z Chinami w tym **formacie** nie jest niczym innym jak realizowaniem naszych praw.

Czynnikiem sprzyjającym rozwojowi stosunków polsko-chińskich jest zainteresowanie Chin Europą, generalnie rzecz biorąc państwami europejskimi. Czynnikiem sprzyjającym jest również inicjatywa „Pas i Szlak”. W obie te koncepcje, w oba te **formaty myślenia chińskiego** o polityce zagranicznej Polska doskonale się wpisuje.

Czy będzie pan kontynuował zapoczątkowany przez pana ministra Sikorskiego **format spotkań Polska - Niemcy - Rosja**, czyli tzw. Trójkąta Kaliningradzkiego?

Na marginesie spotkania Rady ds. Zagranicznych odbędzie się doroczne spotkanie ministrów spraw zagranicznych partnerstwa wschodniego w **formacie dwadzieścia osiem plus sześć**.

Królestwo Norwegii jest bardzo silnie zaangażowane we współpracę w **formacie nordyckim i bałtyckim**, ale docenia takie **formaty** jak chociażby **V4, Grupa Wyszehradzka**;

Ukraina. Za waszych czasów powstawały **formaty, format normandzki, format miński**. Gdzie byliście? Przecież skrytykował was nawet poprzedni minister, minister Schetyna, który domagał się kilka lat temu, że powinniśmy wejść do takiego **formatu**.

Proszę państwa, tutaj tematem była też współpraca LOT-u z lotniskami regionalnymi. Współpracujemy z portami regionalnymi i to nie tylko **w formacie dowożenia pasażerów** do Warszawy, ale także, o czym mówił pan prezes, uruchamiamy oportunistycznie połączenia, które są dla nas rentowne.

III. w trybie art. 151 ust. 1 regulaminu Sejmu z uwzględnieniem art. 3 ust. 2 ustawy z dnia 8 października 2010 r. informację Rady Ministrów o posiedzeniu Rady ds. Zagranicznych (**w formacie ministrów ds. handlu**), które odbędzie się 10 grudnia 2017 r.

Kolejna runda negocjacyjna będzie za 2 tygodnie na posiedzeniu eurogrupy **w formacie inkluzywnym**, czyli znowu z naszym udziałem.

Próbę zdefiniowania nowego *formatu* ułatwić może odszukanie jego synonimów tekstowych. A tych w korpusach nie brakuje, np.:

Rada będzie obradowała tym razem w trzech różnych **formatach**. Po pierwsze to będzie **spotkanie** regularne ministrów spraw zagranicznych. Polskę będzie reprezentował Witold Waszczykowski. Drugą odsłoną będzie robocze **spotkanie** szefów MSZ i szefów MON, a kolejnym **formatem** będzie śniadanie robocze szefów obrony narodowej państw członkowskich UE.

rozwiązaniem konfliktu zajmuje się **Grupa Mińska OBWE**, gdzie główną rolę odgrywają 3 państwa. Jest to Federacja Rosyjska, Francja i Stany Zjednoczone. W ostatnim okresie bardziej aktywna jest tutaj Federacja Rosyjska, trochę mniej Francja, a Stany Zjednoczone nie wykazują jakiejś większej aktywności. Oczywiście obecność Moskwy w tym **formacie** powoduje, że Azerbejdżan musi dosyć, powiedziałbym, balansować, jeśli chodzi o ambicje międzynarodowe.

Chodzi o współpracę w ramach **Grupy Wyszehradzkiej** – przynajmniej z punktu widzenia politycznego jest istotne, żeby także w NATO wyeksponować tę współpracę, a także szczególnie zaznaczaną w ostatnich 2 latach współpracę w ramach „dziewiątki bukaresztańskiej”, owych spotkań prezydentów zainicjowanych 2 lata temu. W tym roku też zapewne prezydenci spotkają się w tym **formacie**. Jeśli chodzi o Brukselę – sprawdziliśmy to już jakiś czas temu i będziemy to kontynuować – to spotkania w ramach tej **grupy** pozwalają wyeksponować interesy i wagę Europy Środkowej wobec innych sojuszników, szczególnie Francji czy Hiszpanii.

Po drugie, jest to współpraca w zakresie politycznym zarówno w **wymiarze** dwustronnym, jak i wielostronnym. Zadaniem będzie tutaj utrzymanie wysokiej dynamiki wizyt na szczeblu prezydentów, premierów, członków gabinetu oraz parlamentu. I ważna będzie współpraca w tym **formacie** multilateralnym, poczynając od **formatu** bałtyckiego, przez Dziewiątkę bukaresztańską, po inicjatywę Trójmorza.

A jeżeli jest tyle problemów, jak twierdzicie, czy nie warto powrócić do sprawdzonych **formatów i grup roboczych** do spraw trudnych? Kiedyś działała **grupa** polsko-rosyjska, być może trzeba stworzyć nową, polsko-ukraińską, oddać rozwiązywanie sporów dotyczących interpretacji historii i postaci historycznych profesorom i historykom.

A zatem jesteśmy aktywni i w **formacie** weimarskim i dwustronnym. Miałem serię spotkań z przedstawicielami misji OBWE na Ukrainie oraz trójstronnej **grupy** kontaktowej, która jest podstawą **procesu** mińskiego. Były bodajże co najmniej trzykrotne spotkania ministerialne **na szczeblu** wiceministerialnym.

Od 2014 r. przedstawiciele trilateralnej **grupy** kontaktowej ds. uregulowania konfliktu na Donbasie regularnie spotykali się w Mińsku. W tym roku '**format** miński' zostanie zakończony.

W naszym interesie jest też wzmocnienie **formatów** współpracy wielostronnej w regionie Morza Bałtyckiego, takich jak **wymiar** północny i Strategia Unii Europejskiej dla regionu Morza Bałtyckiego, Rada Państw Morza Bałtyckiego, Rada Arktyczna, której Finlandia przewodzi do roku 2019, czy **formuły** dialogu państw Grupy Wyszehradzkiej z krajami nordyckimi, bałtyckimi na szczeblu ministrów spraw zagranicznych, zainicjowanego przez Polskę w 2013 r.

Proces chartumski to **format** współpracy Unii Europejskiej z Unią Afrykańską.

Jak widać, najczęściej wśród tekstowych synonimów *formatu* pojawia się rzeczownik: *grupa*, rzadziej *szczebel*, *proces*, *formuła*, *wymiar*, *spotkanie*.

Na uwagę zasługuje również towarzyszący *formatowi* metatekst, który może dostarczyć odpowiedzi o status najczęstszych połączeń, a może także o status samego *formatu*:

Panie Ministrze, Panie Profesorze, jest kilka kwestii, które mnie interesują. Mianowicie jedna jest taka. Czy sprawa tych budzących kontrowersje zmian w edukacji była przedmiotem rozważań Rady Dialogu Społecznego, czy to **w formacie, nazwijmy to**, trój-, czy dwustronnym?

Bardzo ważnym elementem naszej współpracy na pewno jest kwestia wymiaru północnoatlantyckiego. Te dwa państwa członkowskie NATO od dłuższego czasu prowadzą też dialog strategiczny. Dotychczas odbyły się 2 spotkania **w tzw. formacie** małej kwadrygi z udziałem przedstawicieli ministerstw spraw zagranicznych i ministerstw obrony obydwu państw.

Rozmowy **w tzw. formacie** normandzkim (Rosja, Ukraina, Niemcy, Francja) prowadzone od lat 2014–2015 nie przyniosły poprawy sytuacji w Donbasie i nie zbliżyły stron konfliktu do pokoju.

Spotkanie doradców liderów „**formatu normandzkiego**” w Berlinie w sprawie uregulowania konfliktu w Donbasie trwało niemal 9 godzin i nie przyniosło wymiernego rezultatu.

Wstępna hipoteza, iż prawdopodobnie mamy do czynienia z wykształcaniem się nowego znaczenia rzeczownika *format*, pierwsze potwierdzenie znalazła dzięki analizie kolokacji. Drugim, ostatecznym (?), weryfikatorem prawdziwości tezy mogą stać się zaobserwowane powyżej modulujące wypowiedź operatory *tzw. czy nazwijmy to*, a także ujmowanie w cudzysłów wyrażen zawierających słowo *format*.

Format w nowym sensie, inaczej niż w pozostałych (notowanych przez słowniki ogólnej polszczyzny), wykazuje większą łączliwość tak lewo-, jak i prawostronną, czego dowodem są, nie zawsze fortunate, kolokacje typu: *w formacie stolika, format dowożenia pasażerów czy śniadanie w formacie*. Niejednoznaczność (mglistość, nieostrość) pojęcia *format* i statusu związków, które tworzy, zwłaszcza w połączeniu z przymiotnikami odtoponimicznymi, przekłada się także na niepewność statusu onimicznego wyrażen. Świadectwem tego są zapisy wariantywne typu: *Format Normandzki / format normandzki*.

Jak pokazują dane korpusowe dopełniające dane słownikowe, jesteśmy świadkami wykształcania się nowego znaczenia znanego od XVI wieku rzeczownika *format*. Przeprowadzona analiza, łącząca ujęcie właściwe semantyce historycznej i badania korpusowe, potwierdza słuszność tezy Górskiego, Król i Edera, że: „Elektroniczny korpus pozwala [...] na znacznie więcej, bo daje wgląd w przebieg i dynamikę zmiany, dostarczając przykładów tekstowych w takiej liczbie, o jakiej nie mógł marzyć historyk języka pracujący z drukowanymi wydaniem dawnych tekstów” (Górski i in., 2019, s. 16). Analiza kontekstów, w których nowy, nienotowany w słownikach języka polskiego, sens rzeczownika *format* się pojawia, pozwala go wstępnie definiować jako ‘strukturę zorganizowaną wedle pewnych reguł’ – a więc blisko drugorzędnego sensu *formuły* ‘organizacja, struktura’ (ISJP 1, s. 416), a zakres jego użycia zawężyć do dyskursu publicystyczno-politycznego. Wobec braku narzędzia do automatycznego rozpoznawania zmian semantycznych w jednostkach funkcjonujących w tekstach⁴, nie tylko konieczne, ale i owocne – okazuje się wykorzystanie metodologii semantyki historycznej uwzględniającej korpusy językowe i narzędzi do automatycznej analizy jednostek.

BIBLIOGRAFIA

SŁOWNIKI WRAZ ZE SKRÓTAMI

ISJP – Bańko, M. (red.). (2000). *Inny słownik języka polskiego*. Wydawnictwo Naukowe PWN.

⁴ Za zwerbalizowanie tej potrzeby dziękuję prof. Michałowi Szczyszkowi (UAM).

- PSWP – Zgólkowa, H. (red.). (1994–2005). *Praktyczny słownik współczesnej polszczyzny*. T. 1–50. Wydawnictwo Kurpisz.
- SJPD – Doroszewski, W. (red.). (1958–1969). *Słownik języka polskiego*. Państwowe Wydawnictwo „Wiedza Powszechna”, <https://doroszewski.pwn.pl>
- SL – Linde, S.B. (1807–1810). *Słownik języka polskiego*. T. 1–6. W drukarni XX. Pijarów.
- SPXVI – Mayenowa, M.R. i in. (red.). (1956-). *Słownik polszczyzny XVI wieku*. Ossolineum – IBL PAN.
- SW – Karłowicz, J., Kryński, A., Niedźwiedzki, W. (1898–1923). *Słownik języka polskiego*. Druk E. Lubowski i S-ka.
- SWil – Zdanowicz, A., Bohusz-Szyska, M., Filipowicz, J. (1861). *Słownik języka polskiego*. Wydany staraniem i kosztem Maurycego Orgelbranda.
- SWJP – Dunaj, B. (red.). (2000). *Słownik współczesnego języka polskiego*. Wydawnictwo Wilga.
- WSJP – Żmigrodzki, P. (red.). (2007-). *Wielki słownik języka polskiego*. Instytut Języka Polskiego PAN, <https://wsjp.pl/>

KORPUSY

- ChronoPress, <https://chronopress.clarin-pl.eu/#/>
- KDP, https://kdp.nlp.ipipan.waw.pl/query_corpus/
- KorBa, https://korba.edu.pl/query_corpus/
- MoncoPL, <http://monco.frazeo.pl/index>
- NKJP, <http://nkjp.pl>
- NOW Corpus (News on the Web), <https://www.english-corpora.org/now/>
- SpokesPL, <http://spokes.clarin-pl.eu>

LITERATURA

- Bomba, R. (2013). Narzędzia cyfrowe jako wyznacznik nowego paradygmatu badań humanistycznych. W: A. Radomski, R. Bomba (red.). *Zwrot cyfrowy w humanistyce* (ss. 56–71). E-naukowiec.
- Chlebda, W. (2016). W stronę optymalizacji słownikowego objaśniania znaczeń. *Stylistyka*, 25, 319–335.
- Górski, R.L., Król, M., Eder, M. (2019). *Zmiana w języku. Studia kwantytatywno-korpusowe*. Instytut Języka Polskiego PAN.
- Graliński, F. (2019). *Against the arrow of time. Theory and practice of mining massive corpora of polish historical texts for linguistic and historical research*. Wydawnictwo Naukowe UAM.
- Ogrodniczuk, M. (2018). Polish Parliamentary Corpus. W: D. Fišer, M. Eskevich, F. de Jong (red.). *Proceedings of the LREC 2018 Workshop ParlaCLARIN: Creating and Using Parliamentary Corpora* (ss. 15–19). European Language Resources Association (ELRA). <https://www.clarin.eu/ParlaCLARIN>.
- Pawłowski, A. (2021). *ChronoPress – Korpus Tekstów Prasowych*. DOI: 10.34616/139101.
- Pęzik, P. (2015). Spokes – a Search and Exploration Service for Conversational Corpus Data. W: *Selected Papers from the CLARIN 2014 Conference, October 24–25, 2014, Soesterberg, The Netherlands, 99–109. Linköping Electronic Conference Proceedings*. Linköping University Electronic Press.
- Pęzik, P. (2020). Budowa i zastosowania korpusu monitorującego MoncoPL. *Forum Lingwistyczne*, 7, 133–150.

- Przepiórkowski, A., Bańko, M., Górski, R.L., Lewandowska-Tomaszczyk, B. (red.). (2012). *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PAN.
- Przybylska, R. (2007). Z problemów słownikowego opisu częstotliwości rzeczownika. W: P. Żmigrodzki, R. Przybylska (red.). *Nowe studia leksykograficzne I* (ss. 69–77). Lexis.
- Przybylska, R. (2020). Kolokacje a analiza semantyczna wyrazu. *LingVaria*, 15, (2), 43–51.
- Sieradzka-Baziur, B. (2017). Semantyka historyczna – zakres badań i ich wyniki. *Annales Universitatis Paedagogicae Cracoviensis. Studia Linguistica*, 12, 226–237.
- Woźniak, E. (2011). Z warsztatu badacza semantyki historycznej – uwagi o metodzie i źródłach. W: B. Dunaj, M. Rak (red.). *Badania historycznojęzykowe. Stan, metodologia, perspektywy* (ss. 71–79). Księgarnia Akademicka.
- Wybrany y Zupelny Kancyonāl Gdanski* [...]. Thomas Jan Schreiber w Gdansku. Gdańsk 1737.

KORPUSY JĘZYKOWE A/I SEMANTYKA HISTORYCZNA – *FORMAT*

Streszczenie

W artykule podjęto temat semantyki historycznej i możliwości, jakie daje zastosowanie właściwej jej metodologii i uwzględnienie danych i narzędzi cyfrowych. Wyzyskanie słowników (SPXVI, SWil, SL, SW, SJPD, USJP, WSJP PAN), korpusów językowych (Odkrywka, Chronopress, SpokesPL, MoncoPL, Narodowy Korpus Języka Polskiego i Korpus Dyskursu Parlamentarnego), konkordancji i statystyki umożliwiło prezentację wykształcania się nowego znaczenia leksemu *format*, jego synonimów tekstowych oraz chronologizację zmian. Wykazano, że dane korpusowe wyprzedzają leksykografię, dzięki czemu mogą być wykorzystywane do śledzenia neosemantyzmów *in statu nascendi*, a tym samym prezentowania dynamiki zmian w semantyce.

Słowa kluczowe: semantyka; leksykologia; leksem; korpusy językowe; historia języka

LANGUAGE CORPORA AND HISTORICAL SEMANTICS: *FORMAT*

Summary

The article addresses historical semantics and the possibilities offered by its methodology, as well as the inclusion of digital data and tools. The utilization of dictionaries (SPXVI, Swil, SL, SW, SJPD, USJP, WSJP PAN), language corpora (Odkrywka, Chronopress, SpokesPL, MoncoPL, Narodowy Korpus Języka Polskiego/National Corpus of Polish, Korpus Dyskursu Parlamentarnego/Corpus of Parliamentary Discourse), concordances and statistics made it possible to present the emergence of a new meaning of the lexeme *format* and its textual synonyms, and the chronology of the changes. It is shown that corpus data predate lexicography, hence the possibility of using such data to track neosemantisms *in statu nascendi*, and thus presenting the dynamics of semantic changes.

Keywords: semantics; lexicology; lexeme; language corpora; language history