

Mirosław Bańko   
Uniwersytet Warszawski, Warszawa  
[m.banko@uw.edu.pl](mailto:m.banko@uw.edu.pl)

## KONTEKSTOWE WAHANIA FREKWENCJI WZGLĘDNEJ W PARACH SYNONIMÓW I WARIANTÓW

**Słowa kluczowe:** kolokacje, synonimia, warianty leksykalne

**Keywords:** collocations, synonyms, lexical variants

W badaniach porównawczych bliskich sobie jednostek słownictwa – czy będą to synonimy, czy warianty leksykalne – różnice zwykle ujawniają się w kolokacjach. Mimo częstego korzystania z kolokacji w badaniach i zastosowaniach praktycznych różnego typu nie wykształcił się standard analizy kolokacyjnej. Raz bierze się pod uwagę określoną liczbę najistotniejszych kolokatów, np. po 10 pierwszych dla każdego z porównywanych wyrazów; innym razem uwzględnia się kolokaty o określonej charakterystyce gramatycznej, np. tylko przymiotniki; jeszcze inny sposób polega na analizie zasadniczo wszystkich kontekstów znalezionych przez kolokator w korpusie językowym, po uporządkowaniu ich tak, aby uwidocznić preferencje w łączliwości leksykalnej badanych wyrazów.

Wybór metody może być uzasadniony celem postępowania, a także cechami swoistymi badanej pary wyrazowej. W niniejszym artykule zwracam uwagę na użyteczność frekwencji względnej analizowanych wyrazów w wybranych kontekstach jako wskaźnika różnic w ich preferencjach leksykalnych. Mówiąc dokładniej, chodzi o to, by frekwencję względną wyrazów A i B, zmierzoną w całym korpusie, porównać z ich frekwencją względną w kontekstach diagnostycznych i w ten sposób dowiedzieć się, które konteksty faworyzują jednostkę A, a które B. Metoda jest prosta

i trudno w niej upatrywać pomysłu oryginalny. Chcę jednak pokazać, że w tak prosty sposób można czasem uzyskać lepsze wyniki niż przez całościowy ogląd kolokacji wyszukanych w korpusie.

Układ artykułu jest następujący. Zacznę od wyjaśnienia, jak rozumiane są kolokacje, i od kilku podstawowych spostrzeżeń na ich temat, które można znaleźć w literaturze przedmiotu. Stąd przejdę do przykładów analizy kolokacyjnej, by w kolejnej części artykułu pokazać, jak wybór kontekstu zmienia frekwencję względną w parze synonimów lub wariantów.

Choć poświęcony głównie kwestii metodologicznej, niniejszy artykuł dotyka też pewnej kwestii teoretycznej, mianowicie statusu wariantów leksykalnych, które różnią się bardziej, niż zwykle przewiduje się dla zjawiska wariacji.

### Kolokacje jako składnik znaczenia słowa

Kolokacje można rozumieć różnie. Mówiąc ogólnie, chodzi o bezpośrednie lub bliskie sąsiedztwo słów, ale szczegóły bywają definiowane wielorako.

Twórca pojęcia, John Rupert Firth<sup>1</sup>, ujmował znaczenie kolokacyjne (*meaning by collocation*) jako klasę typowych kontekstów leksykalnych słowa i przeciwstawiał znaczeniu pojęciowemu. Ukuł on zgrabny i często cytowany slogan „You shall know a word by the company it keeps”<sup>2</sup> (Firth 1957: 11), zainspirował późniejszych badaczy, ale pojęcia kolokacji nie uściślił. Uczeń Firtha, John Sinclair (1991: 170), ujmował kolokację po prostu jako współwystąpienie dwóch lub więcej słów na krótkim odcinku tekstu, co – zauważmy – nie przesądzało ani o frekwencji takich ciągów, ani o stopniu ich idiomatyczności, ani nawet o tym, czy stanowią one syntagmę, czyli całość składniową. Na gruncie lingwistyki korpusowej przyjmuje się węższą definicję: kolokacjami są tylko takie współwystąpienia słów (zwykle ciągi dwuelementowe), których frekwencja jest istotnie wyższa, niżby wynikało z frekwencji ich składników. Identyfikacja tak rozumianych kolokacji jest oparta na testach statystycznych, np. kolokator wbudowany w wyszukiwarkę PELCRA w Narodowym Korpusie Języka Polskiego (NKJP) korzysta z testu chi-kwadrat, korpus internetowy MoncoPL z tzw. współczynnika Dice’a, a baza kolokacji NKJP, o nazwie HASK, pozwala na wybór spośród kilku testów tego rodzaju (Pęzik 2012, 2013, 2020). Należy dodać, że test testowi nierówny, a ich użyteczność zależy od celu postępowania, np. test log-Dice faworyzuje w kontekście słowa o wysokiej frekwencji, podczas gdy MI (*mutual*

1 Pojęcie kolokacji najsilniej wiąże się z brytyjską tradycją badań nad językiem, ale to samo zjawisko, różnie nazywane, opisywali też lingwiści z innych krajów, niektóre przykłady wymienia Hanna Burkhardt (2010).

2 Poznasz słowo po jego towarzystwie.

*information*) pozwala wyszukać konteksty rzadsze i bardziej swoiste dla ośrodka kolokacji (Čermáková b.d.).

Statystyczne ujęcie kolokacji otwiera furtkę dla ciągów o różnym statusie frazeologicznym, czyli niejednakowym stopniu idiomatyczności. W praktycznych zastosowaniach, np. w leksykografii i w dydaktyce języków obcych, zakres kolokacji jednak zawęża się do zwrotów i wyrażeń, których składniki zachowują swoje znaczenie i które w rezultacie są na ogół zrozumiałe dla odbiorcy, ale nieprzewidywalne dla nadawcy (np. nadawca musi wiedzieć, jakim czasownikiem przekazać ogólne znaczenie ‘robić’ przy różnych rzeczownikach zależnych, por. *podjąć decyzję, sprawić niespodziankę, wyrzucić krzywdę, złożyć skargę*). W tradycji polskiej frazeologii kolokacje tak rozumiane odpowiadają przede wszystkim związkom frazeologicznym łączliwym, rzadziej luźnym (Skorupka 1967: 6–8), przy czym spośród tych ostatnich do kolokacji zalicza się tylko syntagmy istotne poznawczo lub nieoczywiste, normowane przez uzus, a nie system języka (Rokicka 2007: 81; Przybylska 2018: 117, 2020: 48). W literaturze przedmiotu podkreśla się często (np. Fontanelle 1998: 191), że kolokacje zajmują rozległy obszar o nieostrych granicach między dwoma biegunami skali, z których jeden stanowią idiomy (w terminologii Skorupki – stałe związki frazeologiczne), drugi zaś luźne połączenia wyrazowe.

Składniki kolokacji nie są równorzędne: jeden, zwany jej ośrodkiem, jest wyróżniony, np. do celów badawczych, pozostałe, tzw. kolokaty<sup>3</sup>, to słowa, które mu towarzyszą. Składników kolokacji można oczekiwać z dającym się obliczyć prawdopodobieństwem (Firth używał terminu *mutual expectancy* (1957: 12–13)), przy czym prawdopodobieństwo to jest nierówne. Na przykład w kolokacji *zgliwiały ser* obecność słowa *ser* jest prawie pewna<sup>4</sup>, podczas gdy użycie słowa *ser* nie pociąga za sobą z równym prawdopodobieństwem użycia słowa *zgliwiały*, które samo w sobie jest rzadkie (w zrównoważonej części NKJP występuje tylko pięć razy, czyli w przybliżeniu raz na 50 milionów słów).

Firth utrzymywał, że na znaczenie słowa składają się typowe przykłady jego użycia w tekstach. W ujęciu późniejszych badaczy znaczenie kolokacyjne obok innych stało się nieodzowną częścią pełnej charakterystyki semantycznej słów, a także jednostek taksonomicznych wyższego rzędu. Na przykład Geoffrey Leech (1981: 9–23) wyróżnił siedem jej elementów (określił je jako *types of meaning* ‘typy znaczenia’), a Renata Przybylska (2003: 182), wyraźnie w duchu jego propozycji, choć bez

3 Ang. *collocate*. W tym samym znaczeniu używa się terminu *kolokant* (np. Przybylska 2007: 70, 2018: 116; inaczej w: Przybylska 2020) i *kolikator* (np. Burkhardt 2010, za Franzem Josefem Hausmannem), ostatnia z tych nazw może być jednak myląca, gdyż odnoszona jest także do narzędzi pozyskujących kolokacje z korpusu, np. w witrynie wyszukiwarki PELCRA w NKJP. Nawiasem mówiąc, termin *kolokacja* też miewa dwojakie odniesienie: do połączeń wyrazowych pewnego typu i do łączliwości wyrazowej jako takiej (w obu znaczeniach używa go Burkhardt 2010).

4 Słowo *ser* nie reprezentuje tu jednak klasy zamkniętej, może być zastąpione np. zdrobnieniem lub nazwą gatunkową sera.

powołania się na autora, wskazała sześć, częściowo innych. Przez zsumowanie i nieznaczne rozwinięcie tych dwóch propozycji daje się utworzyć listę 10 komponentów szeroko rozumianego znaczenia słowa, obejmującą znaczenie desygnacyjne, konotacyjne, stylistyczne, afektywne, kolokacyjne, tematyczne, strukturalne, etymologiczne, gramatyczne i tzw. znaczenie odbite (szczegóły podają Bańko, Svobodová, Rączaszek-Leonardi, Tatjewski 2016: 66–70).

Można się spierać o hierarchię tych „znaczeń”<sup>5</sup>, a w szczególności o to, czy kolokacje są wtórne wobec znaczenia desygnacyjnego, określającego, do czego dane słowo się odnosi. Dla dziecka, które uczy się znaczenia słów swojego języka przez obserwację kontekstów leksykalnych i sytuacyjnych, kolokacje mają charakter prymarny. Nawet jednak jeśli przyjąć perspektywę osoby dorosłej i uznać kolokacje za wtórne, trzeba przyznać, że dzięki nim łatwiej opisać znaczenie desygnacyjne, a w szczególności wyodrębnić znaczenia słów polisemicznych, odpowiadające różnym klasom kontekstów. Kolokacje pomagają też w opisie konotacji słowa i w pewnym stopniu innych elementów jego charakterystyki funkcjonalnej. Zaletą kontekstów jako przedmiotu badań jest to, że łatwo poddają się różnym operacjom: można je automatycznie selekcjonować, sortować, porównywać i hierarchizować według wybranych kryteriów. Pomaga to w ocenie różnic między badanymi wyrazami, takimi jak pary synonimów lub wariantów leksykalnych.

### Jak analizować kolokacje?

Liczba kolokacji wyszukanych automatycznie w korpusie jest tym większa, im częstszy jest ośrodek kolokacji. Na przykład kolokator PELCR-y znajduje 250 kolokacji leksemu *komponent*, występującego w przybliżeniu 15 razy na milion słów, i 505 kolokacji leksemu *składnik*, który występuje około 38 razy na milion słów<sup>6</sup>.

Niezależnie od tego, jak długa jest lista wyników, od pewnego miejsca pojawiają się na niej kolokaty mało charakterystyczne dla badanego wyrazu, np. zaimki, partykuły, czasowniki posiłkowe lub modalne typu *być*, *mieć*, *móc*. Wbrew pozorom znaczące mogą być przyimki: można je dołączyć do ośrodka kolokacji i powstałe w ten sposób wyrażenie uczynić ośrodkiem nowej kwerendy, aby ujawnić kolokaty,

- 
- 5 Terminu tego używam konwencjonalnie, mając na myśli komponenty charakterystyki funkcjonalnej słowa. Nazywanie znaczeniem np. zespołu cech gramatycznych lub stylistycznych może budzić obiekcje.
  - 6 Z niejasnych powodów PELCRA wskazuje wyższą liczbę kolokacji na gorze utworzonej listy niż liczba kolokacji faktycznie znalezionych. Co więcej, podana tam liczba wystąpień ośrodka kolokacji jest myląca, gdyż odnosi się do liczby zawierających go akapitów. W artykule podają statystyki uwzględniające faktyczną liczbę wystąpień słowa i faktyczną liczbę znalezionych kolokacji. Wszystkie kwerendy, jeśli nie wskazano inaczej, zostały wykonane w zrównoważonej części NKJP przy domyślnych ustawieniach wyszukiwarki PELCRA.

które nie sąsiadują bezpośrednio z badanym słowem. Podobnie można wykorzystać spójniki i znaleźć kolokacje, w których badane słowo jest składnikiem konstrukcji współrzędnej. Innym sposobem na znalezienie kolokatów, które nie znajdują się w bezpośrednim sąsiedztwie ośrodka kolokacji, jest zmiana ustawień poprzez zwiększenie kontekstu, w którym kolikator szuka kolokatów. Wówczas jednak może zostać znalezionych wiele słów nierelevantnych.

Wyników kwerend nie można przyjmować bezkrytycznie. Zdarzają się błędy lematyzacji, nazwy własne, których kolikator nie odróżnia od równokształtnych nazw pospolitych, wyrazy obcojęzyczne, kolokaty zawdzięczające swoją obecność na liście popularnym filmom, w których tytułach wystąpiły wraz z ośrodkiem kolokacji – katalog takich niespodzianek jest obszerny. Ważniejsze jednak pozostaje pytanie, które kolokacje uwzględnić w opisie i w jaki sposób zdać sprawę z preferencji kolokacyjnych danego słowa.

Nie wykształcił się standard badawczy w tym zakresie, ponieważ o sposobie postępowania decyduje jego cel i swoistość przedmiotu badań. Jeśli celem jest np. zbadanie, jak w świetle kolokacji rysuje się polski stereotyp blondynki, to miarodajnych danych dostarcza już analiza lewostronnych kolokatów przymiotnikowych leksemu *blondynka*. PELCRA ujawnia, że dotyczą one przeważnie wyglądu takiej osoby, a pierwsza cecha odnosząca się do osobowości pojawia się dopiero pod koniec drugiej dziesiątki wyników i wyrażona jest słowem *głupia*.

Jeśli badanie kolokacji służy porównawczej analizie synonimów lub wariantów leksykalnych, często istotne różnice można wykryć już dzięki porównaniu określonej liczby początkowych kolokatów jako statystycznie najbardziej istotnych. Na przykład na pierwszych 15 pozycjach listy rzeczownikowych kolokatów postpozycyjnych leksemu *miłośnik* występują kolejno:

cieplice, cieplica, koty, kot, motoryzacja, hippika, fantastyka, militaria, muzyka, wędkarstwo, muzyk, kolarstwo, przyroda, ziemia, jednoślad.

Błędy lematyzacji i zapis wyników małymi literami nieco zaciemniają obraz, ale i tak widać, że przedmiotem miłośnictwa mogą być bardzo różne rzeczy, np. miejscowości (tu Cieplice Śląskie) i regiony kraju (tu np. ziemia tarnogórska), zwierzęta, sport, muzyka i hobby. Potwierdzeniem jest dalsza część listy, zawierająca m.in. *astronomię, poezję, kino, starocie i kaktusy*.

Tej samej długości lista analogicznych do wspomnianych kolokatów leksemu *wielbiciel* zawiera podobne wyniki, więc nie pozwala na odpowiedź, czym różni się wielbiciel od miłośnika:

talent, twórczość, jazz, sztuka, muzyk, muzyka, proza, serial, aktorka, kin, kina, kino, koty, kot, taniec.

Ale już lista sporządzona na tych samych zasadach dla leksemu *fan* sugeruje zupełnie inny profil zainteresowania fanów, obejmujący przede wszystkim sport i muzykę popularną<sup>7</sup>:

speedway, club, motoryzacja, futbol, żużel, basket, jazz, rock, blues, muzyka, muzyk, piłka, koszykówka, gotyk, widzów.

Aby ułatwić porównanie kolokatów dwóch badanych leksemów, warto podzielić je na trzy grupy: do jednej należą kolokaty swoiste dla pierwszego leksemu, do drugiej kolokaty swoiste dla drugiego, a do trzeciej kolokaty wspólne dla obu wyrazów. Baza kolokacji HASK (Pęzik 2013), obejmująca kolokacje wygenerowane z NKJP, ułatwia to zadanie, gdyż udostępnia narzędzie o nazwie Kolozaurus, które grupuje osobno kolokaty wspólne i kolokaty swoiste. Podziału tych ostatnich na dwie grupy użytkownik musi dokonać już sam, np. za pomocą filtrowania wyników w arkuszu kalkulacyjnym Excel. Kolozaurus oferuje też poglądowe wykresy, które pokazują, jak kolokacje dzielą się pomiędzy dwa porównywane wyrazy.

Cennych obserwacji może dostarczyć tematyczne pogrupowanie kolokatów. Na przykład opis obrazu mniejszości romskiej w języku czeskim i słowackim Milena Hebal-Jeziarska (2012) poprzedziła podziałem kolokatów pomiędzy takie grupy, jak dyskryminacja, integracja, praca, przestępstwo, społeczeństwo, i wskazaniem, jaka część kolokatów przypada na każdą z tych grup. Efekty pogrupowania autorka określa jako obrazy kolokacyjne (Hebal-Jeziarska 2010), wizerunki kolokacyjne (Hebal-Jeziarska 2012) lub profile kolokacyjne (Hebal-Jeziarska 2013). Ostatni z tych terminów jest tłumaczeniem angielskiej nazwy *collocation profile* (obocznie *collocational profile*), która jednak nie ma wyraźnego odniesienia, gdyż określa różne sposoby porządkowania kolokatów.

W niektórych zastosowaniach, np. w leksykografii, pożądane jest zestawienie możliwie wielu kolokacji, które mogą być interesujące dla adresata. Na użytek *Słownika dobrego stylu* (Bańko 2006), w istocie chyba pierwszego polskiego słownika kolokacji, sporządzono schemat hasła słownikowego, osobno dla każdej z trzech głównych części mowy, obejmujący różne rodzaje kolokacji i prezentujący je w ustalonym porządku, opartym w pierwszej kolejności na kryteriach składniowych, a w drugiej na semantycznych. Schemat ten został rozwinięty i wdrożony w projekcie badawczym APPROVAL, w którym porównywano synonimy i warianty leksykalne, a opisany szczegółowo w książce projektowej (Bańko, Svobodová, Rączaszek-Leonardi,

7 Nie znaczy to, by nie można było mówić ekspresywnie np. o *fanie kaktusów*, sugerując, że miłośnictwo nabiera tu cech zaangażowania właściwego fanom sportu lub gwiazd estrady. Nawiasem mówiąc, o *fanie* – słowie wówczas nowym i modnym – pisał pod koniec XX w. Krzysztof Wróblewski (1991), korzystając głównie z wyników ankiety przeprowadzonej wśród uczniów. Lektura tamtej publikacji uświadamia, jak bardzo poszerzył się zasób danych, w tym tekstów źródłowych dostępnych w badaniach nad językiem.

Tatjewski 2016: 75–78). Inne schematy, ale z tą samą intencją, opracowała Renata Przybylska (2018: 117–121) dla *Wielkiego słownika języka polskiego PAN* (WSJP PAN), który wśród słowników ogólnych polszczyzny jest dotąd jedynym zawierającym obszerne listy kolokacji.

Takie niemal wyczerpujące i uporządkowane zbiory kolokacji – przez Przybylską nazywane polem kolokacyjnym, przeze mnie obrazem kolokacyjnym słowa – są pracochłonne. Mają ponadto tę wadę, że zrównują słowa obecne na różnych miejscach listy kolokatów, a więc bardziej i mniej istotne statystycznie. Na przykład zarówno wśród kolokatów wariantu *dealer*, jak i kolokatów wariantu *diler*, znalezionych przez wyszukiwarkę PELCRA, trzy najwyższe miejsca zajmują wyrazy *autoryzowany*, *narkotykowy* i *narkotyk*, ale ich kolejność, czyli ranga, jest różna, czego w polach (obrazach) kolokacyjnych nie widać. Jak stąd wynika, zestawienie wszystkich interesujących kolokatów w polach (obrazach) kolokacyjnych może wręcz utrudnić uchwycenie różnicy w preferencjach leksykalnych badanych słów. W takich sytuacjach warto sprawdzić, jak względna frekwencja tych słów zmienia się w zależności od kontekstu, czyli w jakiej proporcji znajdują się poszczególne kolokacje w porównaniu z proporcją, w której pozostają ośrodki tych kolokacji. Szczegóły zostaną przedstawione w następnej części artykułu.

### Zmienność frekwencji w zależności od kontekstu

W tabeli 1 podano statystyki dotyczące wariantów *dealer* i *diler*, pozyskane za pomocą wyszukiwarki PELCRA w zrównoważonej części NKJP. Pierwszy wiersz poniżej nagłówka określa liczbę wystąpień każdego z wariantów, bez względu na kontekst leksykalny, oraz ich proporcję, czyli frekwencję względną. W drugim wierszu podano frekwencję tych samych wariantów w bezpośrednim prawostronnym kontekście leksemów *samochód* lub *samochodowy*, a w trzecim podobnie w kontekście leksemów *narkotyk* lub *narkotykowy*<sup>8</sup>.

---

8 Kwerendy miały postać *dealer\*\* samochód\*\** i *dealer\*\* samochodowy\*\**, czyli uwzględniały odmianę wyrazów. Wyniki zsumowano ręcznie, gdyż uzyskana w ten sposób liczba okazuje się wyższa niż pozyskana za pomocą prostszej i teoretycznie równoważnej kwerendy z alternatywą kontekstów: *dealer\*\* samochód\*\*|samochodowy\*\**. Niezależnie od wyboru wariantu i od kontekstu ta ostatnia gubi od kilku do kilkunastu wyników.

	<i>dealer</i> [A]	<i>diler</i> [B]	Frekwencja względna [A/B]
Ogółem	1470	1026	1,43
Przed leksemami <i>samo- chód</i> lub <i>samochodowy</i>	121	54	2,24
Przed leksemami <i>narko- tyk</i> lub <i>narkotykowy</i>	79	151	0,52

Tabela 1. Frekwencja względna leksemów *dealer* i *diler* a ich frekwencja względna w kontekstach diagnostycznych (źródło: NKJP)

Jak widać, preferencje leksykalne badanych wariantów są rozbieżne: kontekst związany ze sprzedażą samochodów w autoryzowanych salonach zwiększa przewagę *dealera* nad *dilerem*, czyli ich frekwencję względną w porównaniu z wartością referencyjną, wskazaną w pierwszym wierszu tabeli, natomiast kontekst związany z rozprowadzaniem narkotyków ma działanie odwrotne: zmniejsza udział *dealera* w porównaniu z *dilerem*, czyli ich frekwencję względną w porównaniu z wartością referencyjną. Ewidentnie wariant obcy jest bardziej prestiżowy od spolszczonego i relatywnie częściej wybierany, gdy chodzi o pracę, która uchodzi za atrakcyjną, niż o zajęcie napiętnowane społecznie.

Nie jest to odosobniony przykład zróżnicowania wariantów graficznych. *Jazz* jest formą bardziej profesjonalną (Bańko 2014), używaną w piśmiennictwie fachowym i w publicystyce, podczas gdy *dżez* trafia się z rzadka w literaturze pięknej. *Chipsy* w opinii badanych osób okazują się smaczniejsze od *czipsów*, a *camping* droższy, ale wygodniejszy od *kempingu* (Bańko, Svobodová, Rączaszek-Leonardi, Tatjewski 2016: 123–124). O ile jednak tu referencja wariantów, mimo różnic w konotacji, jest wspólna, o tyle *dealer* różni się od *dilera* preferowanym odniesieniem, gdyż widoczna jest tendencja, aby nazw tych używać na określenie różnych osób. Podobnie jest z wariantami *baseball* i *bejsbol*: wariant obcy, dwuipółkrotnie częstszy w NKJP, odnosi się przede wszystkim do gry i tylko pojedyncze wystąpienia, nie więcej niż 3% ogółu, dotyczą kija używanego w tej grze, podczas gdy wśród wystąpień wariantu spolszczonego kij, przeważnie jako narzędzie chuliganów, stanowi odniesienie ponad połowy użyc. Przykłady te uzmysławiają, że warianty mogą się różnić bardziej, niż się zwykle zakłada, nie tylko frekwencją, dystrybucją stylistyczną, rozmieszczeniem terytorialnym, żywotnością, prestiżem, statusem normatywnym, czy choćby pochodzeniem społecznym i pcią swoich użytkowników. Różnice między *dealerem* i *dilerem* czy *baseballem* i *bejsbolem*, dotyczące preferowanego odniesienia, stanowią wyzwanie dla sposobu definiowania wariacji, por. w *Encyklopedii językoznawstwa*

ogólne: „Elementy jęz., które traktuje się identycznie pod względem funkcjonalnym, chociaż różnią się formalnie” (KP 1999)<sup>9</sup>.

Dla potwierdzenia tych obserwacji przeprowadziłem takie same obliczenia w zrównoważonej części Korpusu Współczesnego Języka Polskiego (KWJP), obejmującym teksty książkowe i prasowe dekady 2011–2020 (zob. tabela 2), oraz w internetowym, powiększonym co dzień korpusie MoncoPL (zob. tabela 3). W obu korpusach spolszczony wariant jest liczniejszy od obcego, odwrotnie niż w NKJP, co można tłumaczyć stopniowym upowszechnianiem się formy graficznie przyswojonej (skądinąd naturalnym procesem adaptacyjnym). Mimo to preferencje dla poszczególnych wariantów pozostały bez zmian: zarówno w KWJP, jak i w MoncoPL wariant obcy zyskuje w kontekstach samochodowych, a traci w kontekstach narkotykowych, tzn. jego frekwencja względem wariantu spolszczonego, odpowiednio, rośnie albo spada w porównaniu z wartością referencyjną, mierzoną niezależnie od kontekstu.

	<i>dealer</i> [A]	<i>diler</i> [B]	Frekwencja względna [A/B]
Ogółem	137	437	0,31
Przed leksemami <i>samo- chód</i> lub <i>samochodowy</i>	15	10	1,50
Przed leksemami <i>narko- tyk</i> lub <i>narkotykowy</i>	7	37	0,19

Tabela 2. Frekwencja względna leksemów *dealer* i *diler* a ich frekwencja względna w kontekstach diagnostycznych (źródło: KWJP)

	<i>dealer</i> [A]	<i>diler</i> [B]	Frekwencja względna [A/B]
Ogółem	20635	40279	0,51
Przed leksemami <i>samo- chód</i> lub <i>samochodowy</i>	2492	1447	1,72
Przed leksemami <i>narko- tyk</i> lub <i>narkotykowy</i>	723	5262	0,14

Tabela 3. Frekwencja względna leksemów *dealer* i *diler* a ich frekwencja względna w kontekstach diagnostycznych (źródło: MoncoPL, dostęp: 5 V 2025)

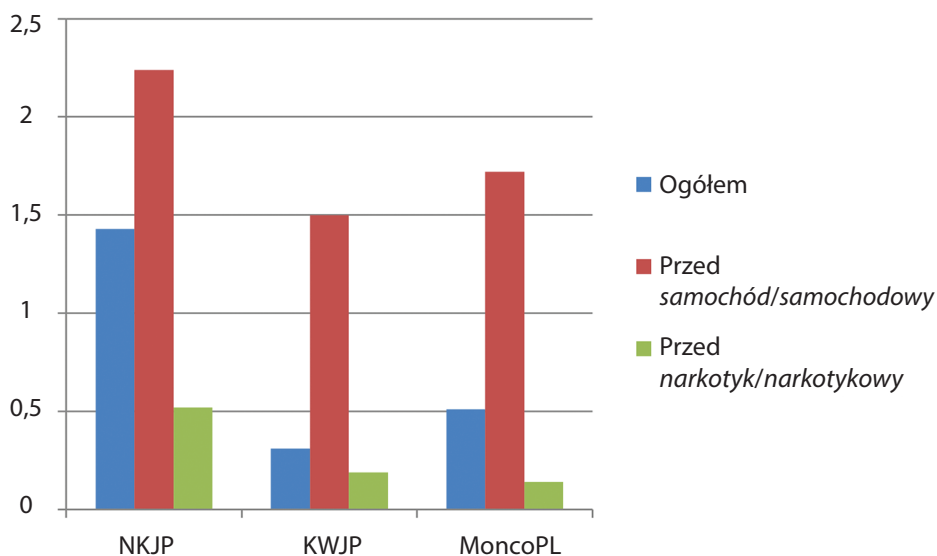
9 Zauważmy przy okazji, że na użytek wariacji leksykalnej definicja ta jest wyraźnie za szeroka. Jeśli wariacja taka ma się czymś różnić od synonimii dokładnej, to warianty mimo różnicy formalnej muszą wykazywać pewne podobieństwo w tym zakresie.

Jak widać, różnica między *dealerem* a *dilerem* rysuje się tak samo w kilku korpusach, mimo że różnią się one zakresem chronologicznym i rodzajem obejmowanych tekstów. Potwierdza ją także analiza obrazów znalezionych dzięki funkcji Grafika wyszukiwarki Google: w wynikach obu kwerend, *dealer* i *diler*, przeważają motywy związane ze sprzedażą narkotyków, ale w obrazach zindeksowanych z formą *diler* jest ich więcej (o przydatności obrazów Google w analizach leksykalnych – zob. Bańko 2013; Kozioł-Chrzanowska 2018).

## Zakończenie

Przedmiotem tego artykułu nie było pozyskiwanie kolokacji z korpusów (np. ocena przydatności poszczególnych testów statystycznych), lecz analiza kolokacji dostarczanych przez narzędzia korpusowe, w szczególności zaś to, jak w kolokacjach przejawia się różnica między synonimami lub wariantami leksykalnymi.

Opisana tu prosta metoda porównywania kontekstowych preferencji synonimów i wariantów ma pewną wadę: konteksty diagnostyczne trzeba wytypować na podstawie innych danych, np. kolokatów lub obrazów Google. Ma też zaletę: pozwala ujawnić wyrazowe sympatie i antypatie wyraźniej niż całościowy ogląd kolokacji, który, jak widzieliśmy, niekiedy zaciera różnice między wyrazami. Wyniki można przedstawić obrazowo w formie wykresu, np. tak jak na wykresie 1.



Wykres 1. Porównanie frekwencji wariantów *dealer* i *diler* w trzech korpusach polszczyzny: ogółem (czyli niezależnie od kontekstu) i w kontekstach diagnostycznych

Choć artykuł dotyczył kwestii metodologicznej – pytania, jak analizować kolokacje – uwypuklił też problem teoretyczny, dotyczący statusu wariantów leksykalnych, których zróżnicowanie jest większe, niż się zwykle zakłada w wypadku wariacji. Należy uznać, że jednostki takie jak *dealer* i *diler* bądź – przywołajmy drugi wspomniany przykład – *baseball* i *bejsbol* to warianty, nie zaś osobne leksemy, ale fakt, że różnią się one preferowanym odniesieniem, sprawia, że stają się interesujące jako wyrazy podlegające tak dużej dywergencji.

## Źródła

- APPROVAL: Adaptation, Perception, and Reception of Verbal Loans, [on-line:] <http://www.approval.uw.edu.pl/> (dostęp: 11 VI 2025).
- HASK: Collocation Databases, [on-line:] [https://pelcra.clarin-pl.eu/hask\\_pl/](https://pelcra.clarin-pl.eu/hask_pl/) (dostęp: 5 V 2025).
- KWJP: Korpus Współczesnego Języka Polskiego, [on-line:] [https://kwjp.ipipan.waw.pl/query\\_corpus/21/](https://kwjp.ipipan.waw.pl/query_corpus/21/) (dostęp: 5 V 2025).
- MONCOPL, MoncoPL, [on-line:] <https://monco-pl.clarin-pl.eu/login> (dostęp: 5 V 2025).
- NKJP: Narodowy Korpus Języka Polskiego, [on-line:] <https://nkjp.pl/> (dostęp: 5 V 2025).
- WSJP PAN: P. Żmigrodzki (red.), *Wielki słownik języka polskiego PAN*, [on-line:] <https://wsjp.pl> (dostęp: 5 V 2025).

## Literatura

- BAŃKO M., 2006, *Słownik dobrego stylu, czyli wyrazy, które się lubią*, Warszawa.
- BAŃKO M., 2013, *Obrazy Google jako źródło informacji lingwistycznej*, [w:] W. Chlebda (red.), *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*, Opole, s. 73–84.
- BAŃKO M., 2014, *Jazz, dżez*, <http://www.approval.uw.edu.pl/jazz> (dostęp: 5 V 2025).
- BAŃKO M., SVOBODOVÁ D., RĄCZASZEK-LEONARDI J., TATJEWSKI M., 2016, *Nie całkiem obce. Zapożyczenia wyrazowe w języku polskim i czeskim*, Warszawa.
- BURKHARDT H., 2010, *Miejsce kolokacji w leksykografii i ich rola w tworzeniu tekstów w języku obcym*, „Acta Universitatis Lodziensis. Kształcenie Polonistyczne Cudzoziemców” 17, s. 177–184.
- ČERMÁKOVÁ A., b.d., *Collocation Analysis*, [w:] M.L. Greenberg (red.), *Encyclopedia of Slavic Languages and Linguistics Online*, <https://referenceworks.brill.com/display/db/eslo> (dostęp: 25 I 2026).
- FIRTH J.R., 1957, *A Synopsis of Linguistic Theory 1930–1955*, [w:] idem, *Studies in Linguistic Analysis*, Oxford, s. 1–32.
- FONTANELLE T., 1998, *Phraseology and the Dictionary*, [w:] A.P. Cowie (red.), *Phraseology: Theory, Analysis, and Applications*, „Oxford Studies in Lexicography and Lexicology”, Oxford, s. 189–207.
- HEBAL-JEZIERSKA M., 2010, *Obrazy kolokacyjne wybranych leksemów, należących do pola semantycznego cudzoziemiec (na materiale korpusu języka czeskiego SYN2006PUB)*, [w:]

- J. Goszczyńska, Z. Greń (red.), *Res slavisticae. Księga dedykowana Profesor Ewie Siatkowskiej*, Warszawa, s. 73–92.
- HEBAL-JEZIERSKA M., 2012, *Wizerunki kolokacyjne mniejszości romskiej w języku czeskim i słowackim*, „Zeszyty Łużyckie” 46, s. 64–81.
- HEBAL-JEZIERSKA M., 2013, *Podstawowe zasady korzystania z korpusów przy badaniu języka*, [w:] W. Chlebda (red.), *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*, Opole, s. 17–30.
- KIERAŚ W., MARCINIAK M., ŁAZIŃSKI M., WOLIŃSKI M., BOJAŁKOWSKA K., EŻŁAKOWSKI W., KOBYLŃSKI Ł., KOMOSIŃSKA D., KRASNOWSKA-KIERAŚ K., RUDOLF M., TOMASZEWSKA A., WOŁOŻYŃ J., ZAWADZKA-PALUEKTAU N., 2025, *Korpus Współczesnego Języka Polskiego. Dekada 2011–2020*, „Język Polski” V, z. 2, s. 5–20, <https://doi.org/10.31286/JP.001055>.
- KOZIOŁ-CHRZANOWSKA E., 2018, *Grafika Google – możliwości i ograniczenia narzędzia jako źródła informacji dla leksykografów redagujących definicje słownikowe*, [w:] M. Bańko, H. Karaś (red.), *Między teorią a praktyką. Metody współczesnej leksykografii*, „Biblioteka Prac Filologicznych (2018)”, t. 1, Warszawa, s. 113–129.
- KP [KAZIMIERZ POLAŃSKI] 1999, *Warianty językowe*, [w:] K. Polański (red.), *Encyklopedia językoznawstwa ogólnego*, wyd. II, Wrocław, s. 630.
- LEECH G., 1981, *Semantics*, Harmondsworth.
- Narodowy Korpus Języka Polskiego*, praca zbiorowa, red. A. Przepiórkowski, M. Bańko, R.L. Górski, B. Lewandowska-Tomaszczyk, Warszawa 2012.
- PĘŻIK P., 2012, *Wyszukiwarka PELCRA dla danych NKJP*, [w:] A. Przepiórkowski, M. Bańko, R.L. Górski, B. Lewandowska-Tomaszczyk (red.), *Narodowy Korpus Języka Polskiego*, praca zbiorowa, Warszawa, s. 253–273.
- PĘŻIK P., 2013, *Paradygmat dystrybucyjny w badaniach frazeologicznych. Powtarzalność, reprodukcja i idiomatyzacja*, [w:] P. Stalmaszczyk (red.), *Metodologie językoznawstwa. Ewolucja języka, ewolucja teorii językoznawczych*, „Łódzkie Studia z Językoznawstwa Angielskiego i Ogólnego”, 1, Łódź, s. 143–160.
- PĘŻIK P., 2020, *Budowa i zastosowania Korpusu monitorującego MoncoPL*, „Forum Lingwistyczne” 7, s. 133–150, <https://doi.org/10.31261/FL.2020.07.11>.
- PRZYBYLSKA R., 2003, *Wstęp do nauki o języku polskim*, Kraków.
- PRZYBYLSKA R., 2007, *Z problemów słownikowego opisu łączliwości rzeczownika*, [w:] P. Źmigrodzki, R. Przybylska (red.), *Nowe studia leksykograficzne*, Kraków, s. 69–77.
- PRZYBYLSKA R., 2018, *Ilustracja materiałowa w WSJP PAN – kolokacje i cytaty*, [w:] P. Źmigrodzki, 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*, Kraków, s. 115–126.
- PRZYBYLSKA R., 2020, *Kolokacje a analiza semantyczna wyrazu*, „LingVaria” XV, nr 2 (30), s. 43–51, <https://doi.org/10.12797/LV.15.2020.30.04>.
- ROKICKA T., 2007, *Status kolokacji w Wielkim słowniku języka polskiego*, [w:] P. Źmigrodzki, R. Przybylska (red.), *Nowe studia leksykograficzne*, Kraków, s. 79–86.
- SINCLAIR J., 1991, *Corpus, Concordance, Collocation*, „Describing English Language”, Oxford.
- SKORUPKA S., 1967, *Słownik frazeologiczny języka polskiego*, t. 1, Warszawa.
- WRÓBLEWSKI K., 1991, *Fan – moda czy językowa konieczność?*, „Poradnik Językowy” z. 3–4, s. 142–145.

*Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*, praca zbiorowa, red. P. Żmigrodzki, M. Bańko, B. Batko-Tokarz, J. Bobrowski, A. Czepakowska, M. Grochowski, R. Przybylska, J. Waniakowa, K. Węgrzynek, Kraków 2018.

### **Contextual Fluctuations in Relative Frequency in Synonym and Variant Pairs**

#### **Abstract**

The author of this article discusses different ways of analysing and presenting collocations in lexical research and shows that in comparative analyses of synonyms and variants, it may be more useful to examine how their relative frequency varies with verbal context than to look at their collocations in their entirety. Some of the variants discussed differ in their preferred reference, which challenges the very definition of variants as functionally identical entities, even if there are differences between them in terms of non-designatory features and sociolinguistic parameters.