

doi: <http://dx.doi.org/10.17651/bptj.74.3>

IWONA KOSEK  
UWM w Olsztynie

SEBASTIAN PRZYBYSZEWSKI  
UWM w Olsztynie

MONIKA CZEREPOWICKA  
UWM w Olsztynie

## **Źródła w tworzeniu elektronicznego słownika paradygmatów frazeologizmów werbalnych<sup>1</sup>**

### Streszczenie

Celem artykułu jest wskazanie typów źródeł i sposobów ich wykorzystania przy opracowywaniu słownika paradygmatów frazeologizmów czasownikowych. Autorzy charakteryzują wykorzystywane źródła, omawiają użyteczność źródeł słownikowych i korpusowych (w zależności od etapu prac nad słownikiem) oraz ich ograniczenia. Szczególną uwagę poświęcają roli introspekcji w tworzeniu słownika gramatycznego.

**Wyrazy kluczowe:** leksykografia, frazeologizmy werbalne, korpus, introspekcja

## **Language data sources in an electronic dictionary of paradigms of verbal phraseological units**

### Abstract

The article discusses types of language data sources and the ways they may be used in the compilation of a dictionary of paradigms of verbal phraseological units. The authors describe the sources they used and discuss the usefulness of lexicographic and corpus sources and their limitations. Special attention is paid to the role of introspection in the process of making a grammatical dictionary.

**Keywords:** lexicography, verbal phraseological units, corpus, introspection

---

<sup>1</sup> Praca naukowa finansowana ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2013/09/B/HS2/01222.

1. Uwagi wprowadzające<sup>2</sup>

Tworząc słownik, korzysta się z trzech typów informacji: własnej kompetencji językowej, wcześniejszych słowników danego typu (jeśli takie istnieją) oraz tekstów dokumentujących zastosowanie (użycie) jednostek (zob. Bańko 2001: 13; Kilgarriff 2013: 77–78)<sup>3</sup>. Wszystkie wymienione elementy można rozumieć jako źródła w leksykografii. W tradycji określenie *źródła* interpretowano zwykle wężiej, odnosząc je do oryginalnych tekstów, na podstawie których opracowywano listę jednostek i ilustracje ich użycia. Oś rozważań leksykografów stanowił przy takim rozumieniu źródeł dobór tekstów i jego konsekwencje dla ustalenia listy opisywanych jednostek (zob. np. Lubaś 1988). Dyskusję nad znaczeniem tekstów źródłowych dla leksykografii i gramatyki, nad stopniem ich wykorzystania w polskich słownikach, a także relacje (różnice) między korpusem a kartoteką opisuje Mirosław Bańko (2001: 25–49, zob. też Bańko 2012), nie będziemy jej zatem w tym miejscu przywoływać. Niewątpliwie o przydatności tekstów źródłowych nie trzeba nikogo przekonywać, zwłaszcza w dobie rozwijającego się językoznawstwa korpusowego, a istniejące zbiory tekstów (przede wszystkim NKJP) stanowią współcześnie istotną, jeśli nie podstawową leksykograficzną bazę materiałową (zob. np. Żmigrodzki 2015).

W prezentowanym artykule jako źródła rozumiemy zarówno zbiory tekstów (oryginalne przykłady użycia), jak i dostępne słowniki. Te drugie można traktować jako metaźródła, kryją się za nimi bowiem konkretne autorskie interpretacje i decyzje. Mówimy zatem w istocie o informacjach z dwu poziomów – „czystego” i „przefiltrowanego”. Chcemy wskazać, jak te informacje są i mogą być przydatne w stworzonym słowniku fleksyjnym frazeologizmów werbalnych (szerzej o słowniku piszemy w p. 2.), a także zwrócić uwagę na pewne ogólniejsze problemy związane z wykorzystaniem różnych typów źródeł.

Stopień przydatności źródeł i ich waga w sposób oczywisty zależą od typu słownika<sup>4</sup>, przewidywanego odbiorcy (zob. Sinclair 1995, Čermák 2003), ale też od

<sup>2</sup> Wkład każdego z autorów w powstanie niniejszego artykułu jest równy (33,33%). Wspólnie opracowano koncepcję i założenia tekstu, zebrano materiał przykładowy, przeprowadzono analizę i nadano pracy ostateczny kształt.

<sup>3</sup> Kilgarriff (2013: 77–78) mówi o trzech sposobach tworzenia słownika, są to: kopiowanie, introspekcja i ogląd danych. Pierwszy, rozumiany jako kwerenda po innych, istniejących już słownikach, nie może się z oczywistych względów ograniczać do samego przepisania rozwiązań, o ile dany słownik ma wnieść jakąś nową jakość do leksykografii. Centralna jest, zdaniem Kilgarriffa, introspekcja, choć i ona ma swoje ograniczenia: badacz, opierając się na swojej kompetencji językowej, może pewne rzeczy sobie uświadamiać, lecz inne mogą ująć jego uwadze. W tym względzie pożyteczny jest ogląd danych (korpus).

<sup>4</sup> Twórca słownika gramatycznego, podobnie jak każdy leksykograf, musi wybrać między postawą preskryptywną i deskryptywną. Współczesna leksykografia odeszła od preskryptywności właściwej XVII- i XVIII-wiecznym słownikom, jak choćby *Dictionnaire de l'Académie française* stanowiącemu wzór dla wielu późniejszych opracowań. Jednakże postawa deskryptywna, dowartościowująca rzeczywiste użycie języka, nie zwalnia całkowicie leksykografa od zadania wskazywania odbiorcy form preferowanych – my również, choć zajmujemy postawę deskryptywną, dokonujemy niekiedy arbitralnych, opartych na naszej kompetencji, decyzji, odrzucając pewne formy, mimo że są reprezentowane w tekstach.

zakładanej roli korpusów. Jak wiadomo, wyróżnia się 3 podejścia w badaniach z ich wykorzystaniem: badania ilustrowane przez korpus (*corpus-illustrated / informed approach*), oparte na korpusie (*corpus-based approach*) i sterowane korpusem (*corpus-driven approach*), zob. np. Tummers, Heylen, Geeraerts 2005, Górski 2012, Grabowski 2015. W tworzonym słowniku prezentujemy podejście drugie, wykorzystujemy zatem korpus nie tylko jako źródło oryginalnych przykładów, ale przede wszystkim do weryfikowania / falsyfikowania hipotez i sprawdzania ich pełności. Trzecie, najbardziej radykalne, podejście w tworzonym słowniku jest badawczo mniej przydatne. Po pierwsze z tego względu, że widzimy frazeologię jako element systemu (pewną grupę jednostek leksykalnych), nie jako zjawisko tekstowe; po drugie – bardzo liczne związki frazeologiczne zostały już wydobyte z tekstów, zarejestrowane, nie opisano natomiast ich paradygmatów i na tym zadaniu się skupiamy; po trzecie wreszcie, jak pokazują dotychczasowe badania (zob. np. Moon 1998; Kilgarriff 2013: 80), najlepiej frekwencyjnie wyodrębniają się z tekstów regularne semantycznie kolokacje, a ich fleksja jest wysoce przewidywalna, inaczej niż jednostek o zidiomatyzowanym znaczeniu. Przedstawione względy decydują o tym, że rolę korpusów w tworzonym słowniku można postrzegać jako drugoplanową: listę jednostek ustalamy głównie na podstawie istniejących słowników i własnej kartoteki, a podstawę opisu paradygmatów stanowi własna kompetencja językowa. W takich prostych stwierdzeniach roli korpusów tekstowych w słowniku gramatycznym zamknąć się jednak nie da, przynajmniej z dwu powodów. Przede wszystkim specyfika jednostek frazeologicznych decyduje o tym, że nie ma całkowitej odpowiedniości między fleksją (paradygmatem) związku a paradygmatem homonimicznego z głównym komponentem leksemu. W przypadku frazeologii zależność gramatyki od semantyki jest bardzo widoczna, wręcz wyostrowana w zestawieniu z leksemami. Nigdy nie można zakładać, że związek będzie miał taki sam paradygmat jak jego główny komponent – homonimia o niczym nie przesądza. Decydujące są własności znaczeniowo-funkcjonalne frazeologizmu i wynikająca z nich jego tekstowa łączliwość. Z tego względu słownik gramatyczny leksemów (w przypadku tworzonego słownika to SGJP oraz opracowanie *Czasownik polski. Odmiana, słownik Saloniego*, 2001) jest w pracy nad paradygmatyką frazeologii niewystarczający. Ponadto, choć oryginalne teksty dla omawianego typu słownika nie są tak istotne jak dla leksykografii ogólnej, niewątpliwie mogą pomóc w rozwiązaniu kilku typów kwestii szczegółowych<sup>5</sup>.

## 2. Typy źródeł i ich wykorzystanie w pracy nad słownikiem fleksyjnym frazeologizmów

Niniejsze refleksje związane są z pracą nad elektronicznym *Słownikiem paradygmatów polskich frazeologizmów czasownikowych* VERBEL (słownik w postaci pliku jest dostępny na stronie [www.uwm.edu.pl/verbel/](http://www.uwm.edu.pl/verbel/); zob. też Czerepowicka,

<sup>5</sup> Na temat możliwości wykorzystania korpusów w badaniu gramatyki por. m.in.: Conrad 2010; Górski 2012.

Kosek, Przybyszewski 2014). Frazeologizmy traktujemy jako jednostki języka (zgodnie z koncepcją A. Bogusławskiego (1976)), spośród nich w słowniku ujmujemy tylko te, które w funkcjonalnej klasyfikacji związków frazeologicznych Andrzeja Marii Lewickiego (1986) określane są jako zwroty (zob. też Lewicki 1976; Lewicki, Pajdzińska 2001). Słownik, w postaci specjalnie do tego celu opracowanej aplikacji Werbosław<sup>6</sup>, podaje wprost paradygmaty około 5000 zwrotów. Punkt wyjścia badań stanowi założenie, że fleksja (zasób form) jednostek frazeologicznych jest oparta na fleksji ich komponentów i grupy syntaktycznej, której schemat realizują (Lewicki, Pajdzińska, Rejakowa 1987: 20). Zwroty odmieniają się więc tak, jak ich główny czasownikowy komponent (zestaw kategorii fleksyjnych czasownika przyjmujemy za Zygmuntem Salonim, zob. Saloni 2001; SGJP). Ich paradygmat nie jest jednak prostym odbiciem paradygmatu podstawowego składnika. Nieregularne znaczeniowo frazeologizmy werbalne, jak można wnioskować z dotychczasowych badań, mają więcej ograniczeń aniżeli czasowniki, na których są oparte (zob. Bąba 1998; Kosek 2013; Czerepowicka, Kosek, Przybyszewski 2014; Przybyszewski 2015). Podstawowe zadanie przy opisie paradygmatu zwrotu polega zatem na określeniu, czy podlega on ograniczeniom fleksyjnym, a jeśli tak, to w jakim zakresie.

Prace nad omawianym słownikiem można podzielić na trzy etapy: 1) zbieranie materiału i ustalanie postaci jednostek; 2) określanie paradygmatów zgromadzonych zwrotów; 3) wprowadzanie paradygmatów do aplikacji Werbosław. Źródła były przez nas wykorzystywane na dwóch pierwszych etapach.

### 2.1. Gromadzenie materiału i ustalanie postaci jednostek

Źródłami pomocnymi podczas pierwszego etapu prac były przede wszystkim słowniki ogólne języka polskiego (ISJP, SJPD, SJPSz, SWJP, USJP, WSJP) oraz słowniki frazeologiczne (PSF, SFJP, SFWP, WSF, WSFJP). Wykorzystaliśmy głównie te opracowania, które konsekwentnie posługują się zbliżonymi do przyjętych przez nas kryteriami wyodrębniania jednostek nieciągłych, czyli ISJP i WSJP. Wiele zwrotów zawartych w słownikach stosujących inne definicje i odbiegające od przyjętych przez nas metod wyróżniania frazeologizmów okazywało się, z naszej perspektywy, połączeniem czasowników i różnego typu fraz niewerbalnych. Spośród słowników frazeologicznych w największym stopniu korzystaliśmy z tzw. słowników poznańskich, tj. PSF i SFWP. Natomiast słownik Stanisława Skorupki (SFJP), ze względu na bardzo szerokie rozumienie frazeologii – jako nauki o połączeniach wyrazów, traktowany był przez nas jako źródło uzupełniające. Sięgaliśmy również do sond słownikowych (Bogusławski, Garnysz-Kozłowska 1979; Bogusławski, Wawrzyńczyk 1993; Bogusławski, Danielewiczowa 2005). To właśnie z sond zaczerpnęliśmy takie zwroty,

<sup>6</sup> Ścisłe mówiąc, Werbosław jest rozbudowaną wersją aplikacji Toposław, powstałej przy tworzeniu SAWY – Słownika elektronicznego nazewnictwa Warszawy (Marciniak i in. 2009), i wykorzystanej także do budowy SEJF-u – Słownika elektronicznego jednostek frazeologicznych (Czerepowicka 2014). Trzeba jednak zaznaczyć, że program przeszedł znaczne zmiany, uwzględniające specyfikę i poziom skomplikowania polskiej koniugacji, które w konsekwencji umożliwiły kodowanie frazeologizmów czasownikowych.

jak np.: {*ktos* / *coś*} *bije rekordy {czegoś}*; {*coś*} *umiera śmiercią naturalną*, {*coś*} *daje głos* ('szczeka'). Dodatkowym źródłem frazeologizmów były opracowania naukowe. Stanowiły one źródło takich jednostek, jak np. {*coś*} *pada na podatny grunt* czy {*coś*} *przyprawia {kogoś} o zawrót głowy* (Liberek 2006). Ponadto własne kwerendy tekstowe pozwoliły znaleźć jednostki, które nie są dotychczas odnotowane w słownikach. To zwroty zasłyszane, z literatury, prasy czy Internetu. Należą do nich np. {*ktos*} *nie urodził się wczoraj*, {*ktos*} *spuści kogoś na drzewo / po brzytwie*, {*ktos*} *nie ma w wszystkich w domu*, {*ktos*} *ma jakiś ból głowy {z czymś}*, {*ktos*} *czarno {coś} widzi*, {*ktos*} *ugotuje {kogoś} na twardo*, {*ktos*} *umoczy dzioba*.

Korzystając na etapie gromadzenia jednostek z różnych źródeł leksykograficznych, stykamy się, jak łatwo przewidzieć, z różnicami co do postaci i liczby odnotowywanych frazeologizmów. Problem rozbieżności interpretacyjnych (bo tego obrazem są odmienne postaci związków w różnych słownikach) można uznać za typowy dla wielości źródeł, nie do uniknięcia nawet przy podobnym czy tożsamym spojrzeniu metodologicznym na dane zjawisko, w przypadku jednostek wielowrazowych występującej z całą ostrością. Przykładowe rozbieżności co do postaci jednostek podano w poniższej tabeli. Uwzględniono w niej cztery słowniki ogólne i dla porównania słownik frazeologiczny wydawnictwa PWN – WSF:

SJPSz	USJP	ISJP	WSJP	WSF
<i>iść, pójść pod młotek (licytacyjny)</i>	<i>coś idzie, poszło pod (licytacyjny) młotek</i>	'Mówimy, że coś idzie lub poszło <b>pod młotek</b> albo że coś oddano <b>pod młotek</b> , jeśli jest to wystawione na licytacji lub zostało sprzedane na licytacji'; FPS + CZ	<i>coś idzie pod młotek; coś poszło pod młotek</i>	<i>coś idzie, poszło pod (licytacyjny) młotek</i>
<i>pleść trzy po trzy</i>	<i>bredzić, mówić, pot. klepać, pleść itp. trzy po trzy</i>	'Jeśli ktoś plecie, gada, mówi itp. <b>trzy po trzy</b> , to mówi bez sensu'; FPS + CZ	<i>trzy po trzy</i>	<i>bredzić, mówić, klepać, pleść itp. trzy po trzy</i>
<i>mówić na wiatr, rzucać słowa na wiatr</i>	<i>mówić na wiatr, rzucać słowa na wiatr</i>	'Jeśli ktoś nie <b>mówi na wiatr</b> lub nie <b>rzuca słów na wiatr</b> , to nie składa obietnic, których potem nie dotrzymuje'; FCZ NDK ZWYKLE NEG	<i>ktoś rzuci słowa na wiatr; ktoś rzuca słowa na wiatr</i> <i>rzad. ktoś mówi na wiatr</i>	<i>nie mówić na wiatr, nie rzucać słów na wiatr</i>

Tab. 1. Rozbieżności w zapisie frazeologizmów w wybranych słownikach

Przypadków podobnych różnic jest znacznie więcej. Spójrzmy jeszcze na dwa przykłady takich rozbieżności. W SJPSz, USJP i WSF ciąg *patrzeć na kogoś jak na raroga* jest odnotowany jako frazeologizm czasownikowy. Z kolei redaktorzy ISJP i WSJP wymienionego ciągu nie uznają za frazeologizm, lecz tylko za jedno z regularnych znaczeń leksemu *raróg*, por. definicję w ISJP: 'jeśli ktoś patrzy na kogoś jak na **raroga**, uważa kogoś za **raroga** itp., to patrzy na niego jak na jakąś osobliwość, podejrzliwie i niechętnie'. W WSJP na pięć przykładów ilustrujących omawiane zna-

czenie rzeczownika *raróg* tylko jeden to kolokacja z czasownikiem *patrzeć*. Inny typowy przykład opisywanych rozbieżności to ciąg *co koń wyskoczy*, który w SJPSz, USJP i WSF traktowany jest jako fraza czasownikowa z kilkoma wariantami werbalnymi (*gnać / jechać / pędzić co koń wyskoczy*), zaś w ISJP i WSJP opisywany ciąg odnotowuje się jako frazę przysłówkową.

Prezentowane powyżej przykłady ilustrują pewne ogólne tendencje w notacji nieciągłych jednostek werbalnych w poszczególnych słownikach. O ile ISJP oraz WSJP raczej „skracają” postać jednostki, redukując ją do postaci frazy przysłówkowej, o tyle słowniki tradycyjne przejawiają tendencje rozszerzające, jako frazeologizmy werbalne traktując frazy przysłówkowe czy rzeczownikowe. Takie rozbieżności można wytłumaczyć różnicami metodologicznymi (zwłaszcza uwzględnianiem pojęcia klasy otwartej / zamkniętej w ustalaniu granic jednostek w ISJP i WSJP), ale także celami i założeniami leksykografów. Słowniki tradycyjne oraz nowsze wpisujące się w tradycyjną metodologię leksykograficzną mniejszą wagę przywiązują do postaci jednostki nieciągłej i jej gramatycznej charakterystyki, a zasadniczym ich celem jest podanie użytkownikowi typowego sposobu użycia (stąd często pojawiające się w opisywanych przypadkach ciągi bliskoznaczników czasownikowych, np. *brać, czytać, kupować itp. jak leci*). Z kolei w ISJP i WSJP znacznie więcej uwagi poświęcono zdyscyplinowanemu wyodrębnieniu jednostek nieciągłych i ich gramatycznej charakterystyce, a typowe użycia zostały omówione w definicjach i / lub zilustrowane przykładami. Na marginesie warto odnotować, że również między ISJP i WSJP nie ma pełnej zgodności co do postaci frazeologizmów. Wątpliwości „zwrot czy wyrażenie?” ISJP wydaje się rozstrzygać, traktując formy czasowników jako zewnętrzne otoczenie jednostki, por. np. *ktoś nie powiedział ostatniego słowa* (WSJP) vs. *ostatnie słowo* (ISJP), *ktoś nie da złamanego grosza* (WSJP) vs. *złamany grosz* (ISJP), *ktoś nie pamięta o bożym świecie* (WSJP) vs. *boży świat* (ISJP). Podane przykłady rozbieżności potwierdzają, że jednostki nie są dane, trzeba je wydobywać z tekstów, oraz że ostateczna interpretacja należy do leksykografa, podobnie jak interpretacja co do użyteczności danego słownika jako źródła.

Na etapie gromadzenia jednostek pomocne były też dane korpusowe (zwykle z NKJP). Zgodnie z podejściem *corpus-based*, kwerenda korpusowa służyła weryfikacji wstępnych intuicji i hipotez co do postaci jednostki. I tak na przykład dane z NKJP spowodowały modyfikację pierwotnie zakładanej postaci jednostki *{ktoś} ma niezły ból głowy {z czymś}*. Ponieważ teksty wskazują na wymiennność elementu *niezły* w obrębie klasy otwartej (*spory, niemały, niezły, duży, wielki...*), a wymiennność ta nie narusza znaczenia i struktury frazeologizmu, ostatecznie zanotowaliśmy jednostkę *{ktoś} ma {jakiś} ból głowy {z czymś}*.

Sytuację przeciwną ilustruje odnajdywany w słownikach ciąg *{ktoś} koszulę / majtki w zębach nosił*. Korpus, poza postacią notowaną w słownikach, dostarcza przykładów ze składnikiem *pieluchę / pieluchy*:

(1) Nie ma się co obruszać. Po prostu Jagna wtedy pieluchę w zębach nosiła, to i nie pamięta. (NKJP – Usenet – pl.soc.dzieci)

(2) Niestety jako młodzieniec, który w tychże latach (właściwie to za Gierka) stawał dopiero pierwsze kroczki, albo jak wolą inni, pieluchy w zębach nosił, nie mogę ocenić na ile prawdziwe jest to przybliżenie. (NKJP – Usenet – pl.rec.ksiazki)

Nieuzasadnione byłoby jednak na podstawie zaledwie kilku przykładów z forów internetowych, nieznajdujących potwierdzenia w intuicji leksykografa, uznać dany ciąg za kolejny wariant nominalny. Modyfikację zwrotu można w wielu przypadkach zinterpretować jako spontaniczną, jednokrotną innowację nadawcy, mającą podkreślić znaczenie zwrotu, bądź jako nieświadomą kontaminację czy inne błędne przytoczenie zwrotu. W tym wypadku mamy do czynienia ze stosunkowo częstą i kłopotliwą dla leksykografa sytuacją, mianowicie z niedoborem tekstów, ale też z problemem interpretacji źródeł. Problem ten rysuje się ze szczególną ostrością w przykładach znajdujących z pomocą wyszukiwarek internetowych (np. Google, Yahoo).

Niewątpliwą zaletą dużego korpusu tekstów jest to, że daje on leksykografowi pewien zbiór poświadczeń, mogący zawierać takie warianty lub formy analizowanych jednostek, których badacz – odwołując się do swojej intuicji – nie brał wcześniej pod uwagę (Hanks 1990: 40; Biber 2015: 163). W tym sensie korpus wspomaga introspekcję, zapobiegając popadnięciu badacza w pewien schematyzm myślenia. W omawianych wyżej wypadkach dane korpusowe nie przesądzały o decyzjach dotyczących kształtu frazeologizmu, ale były ważnym argumentem, który należało wziąć pod uwagę (Hanks 1990: 32; Bańko 2012: 289). Ostatecznym weryfikatorem postaci jednostki nieciągłej pozostawała kompetencja językowa samych leksykografów.

## 2.2. Ustalanie paradygmatów frazeologizmów werbalnych

Drugi etap pracy nad słownikiem to ustalanie paradygmatów wyodrębnionych nieciągłych jednostek werbalnych. Tu przydatność obu analizowanych typów źródeł jest ograniczona kilkoma względami: korpusów – przede wszystkim przyjmowanym przez nas stanowiskiem nietekstowym, słowników – ubogą na ogół informacją fleksyjną o związkach frazeologicznych. Leksykografia starsza informacji gramatycznej o frazeologizmach nie podaje, co wynika m.in. z myślenia o nich w kategoriach wtórnych elementów systemu lub w kategoriach łączliwości wyrazów, ale też z koncentracji na opisie znaczenia związków. Słowniki nowsze, z których przynajmniej część traktuje frazeologizmy jako jednostki leksykalne i wobec tego opisuje je analogicznie do leksemów, dają pewne wskazówki co do odmiany. Konsekwentnie czynią to ISJP i WSJP (a także sondy słownikowe), choć w różny sposób. WSJP, podając informacje o odmianie zwrotu, odsyła zazwyczaj do jego głównego komponentu, przy którym z kolei znaleźć można paradygmat odtworzony za SGJP. Przy każdym zwrocie jest też informacja o przechodniości oraz o ewentualnym odpowiedniku aspektowym. Ponadto niekiedy dodatkowo zamieszczono informację o ograniczeniach w odmianie lub użyciu form, zob. np. *ktoś nie jest z cukru* – „zwykle cz. ter.”. ISJP z kolei nie odsyła wprost do paradygmatu czasownika, ale przy części związków podaje obok

informacji gramatycznych wskazówki co do ograniczeń fleksyjnych. Opis tych samych zwrotów może się w obu słownikach różnić. I tak na przykład jednostka {*coś*} *piechotą nie chodzi* w ISJP została opatrzona komentarzem: FCZ NDK TYLKO TER, zaś w WSJP podano: „fraza czasownikowa ndk, npch; odmienny: *nie chodzić*”. Z kolei zwrot {*ktoś*} *utopi {kogoś} w łyżce wody* w ISJP skomentowano następująco: FCZ DK ZWYKLE WAR / BEZOK, zaś w WSJP: „fraza czasownikowa dk, npch; odmienny: *utopiłby*. Spotykane również użycia formy czasu przyszłego: *ktoś utopi kogoś w łyżce wody*”. Tego typu informacje są dla autorów słownika fleksyjnego przydatne, jednocześnie jednak niektóre z nich dalekie od celu, jaki sobie wyznaczają. Słownik paradygmatów musi bowiem podawać formy wprost, innymi słowy – przynajmniej przy określonej koncepcji takiego słownika – powinna być w nim wyraźnie postawiona granica między ‘jest’ a ‘nie ma’, explicite powinien on stwierdzać, czy dana forma jest dopuszczalna w paradygmacie czy też – pomimo że dałoby się ją utworzyć (na przykład komponent czasownikowy daje taką możliwość) – jest ona wykluczona, jej sensowne użycie nie jest możliwe (np. z powodu semantycznych własności zwrotu). Dla autorów i użytkowników to informacja zasadnicza.

Określając paradygmat, posługujemy się zatem głównie własną kompetencją językową (introspekcją), porównując ewentualnie przykłady, w których dostrzegamy ograniczenia fleksyjne, z materiałem słownikowym i korpusowym. Korpus jest użyteczny na tym etapie prac, jak wspomniano we wstępie, do weryfikacji hipotez, także pod względem ich pełności (*corpus-based approach*)<sup>7</sup>. Na pewno, przeszukując bazy tekstów, nie można oczekiwać potwierdzenia użycia każdej formy, dla frazeologii werbalnej sytuacja typowa jest bowiem taka, że korpusy (także Internet) nie poświadczają wielu (żeby nie powiedzieć – większości) form z paradygmatu danego zwrotu. I tak na przykład dla zwrotu {*ktoś*} *kupuje kota w worku* w słowniku przyjmujemy pełny paradygmat (zgodny z paradygmatem czasownika *kupować*). Kwerenda w NKJP dostarcza w podkorpusie zrównoważonym poświadczeń dla jedynie 11 form (w tym dla kilku po zaledwie jednym przykładzie). Nie znajdujemy w nim zatem między innymi przykładów użycia omawianego zwrotu w czasie przyszłym i trybie przypuszczającym, 2. os. lm. trybu rozkazującego, zaś czas przeszły jest reprezentowany jedynie przez formę 3 os. lm. rodz. mos. Z kolei cały korpus NKJP zawiera przykłady użycia zwrotu we wszystkich formach czasu teraźniejszego i trybu rozkazującego oraz po jednej formie (3 os. lp. rodz. męs.) dla czasu przyszłego i trybu warunkowego. Takie wyniki w żaden sposób nie przesądzają o pełności / defektywności paradygmatu.

Jako przykład sytuacji, w której korpus jest pomocny w weryfikowaniu hipotez i rozstrzygnięciu wątpliwości, „podpowiadając” rozwiązanie przynajmniej co do niektórych form paradygmatu, może posłużyć przywoływany już wyżej zwrot {*ktoś*} *ko-*

<sup>7</sup> Oczywiście posłużyć może on także do wskazania form i kontekstów typowych, częstych, a więc do uzyskania pewnych informacji pragmatycznych. Informacje takie są jednak w tworzonej słowniku zdecydowanie drugoplanowe, np. danych o frekwencji nie podajemy w sposób systematyczny, niekiedy zamieszczamy jedynie uwagę w komentarzach.



*szulę w zębach nosił* ('ktoś był młody w czasach, o których mowa'), który według WSJP ma paradygmat defektywny – nie jest używany w czasie teraźniejszym i przyszłym. Kwerenda korpusowa dostarcza przekonujących przykładów użycia opisywanego frazeologizmu w czasie teraźniejszym, których zgodnie z informacjami zawartymi w WSJP i ISJP nie należałoby się spodziewać:

(3) Przeżegnajże się lewą ręką, dziewucho! A cóż jabym u Pazderskich robiła? Stara użera się od rana do nocy, czeladzie porozpędzali, same chłopaki, sama koprowina, co jeszcze w zębach koszule nosi. ([https://pl.wikisource.org/wiki/Pod\\_prawem/I](https://pl.wikisource.org/wiki/Pod_prawem/I))

(4) Spać, smyki! – zawołał Pantera ze swego legowiska. – Koszulę to jeszcze w zębach nosi, a już spiskuje. (NKJP; M. Rodziewiczówna, *Lato leśnych ludzi*)

Formy te nie wchodzą w konflikt z semantyką jednostki, podobnie jak formy bezokolicznika:

(5) O, widzicie ją jaka śmiała! A zdawać by się mogło, że powinna jeszcze koszulę w zębach nosić? ([jaswily.pl/gmina/pliki/wesele.doc](http://jaswily.pl/gmina/pliki/wesele.doc))

(6) Przestałem już nosić koszulę w zębach i dostałem swój pierwszy spinning. (<https://www.scribd.com/document/357862607/Ota-Pavel-S-mierc-pie-knych-saren>)

A zatem paradygmat interesującej nas jednostki nie ogranicza się do form czasu przeszłego, chociaż one zdecydowanie przeważają. Akceptowalne są, choć przez użytkowników używane sporadycznie, formy czasu teraźniejszego i bezokolicznik. Brak natomiast czasu przyszłego, a także form trybu rozkazującego, przypuszczającego, imiesłowu i bezosobnika (tych informacji słowniki już nie podają).

Inny przykład tej samej sytuacji to frazeologizm *{cos} spalilo na panewce*, który ma według Andrzeja Markowskiego tylko czas przeszły (Markowski 2005: 235). W opracowaniu *Kultura języka polskiego* błędną formę zilustrowano następującym cytatem:

(7) \*Tym samym jeden z celów wprowadzenia systemu, czyli usprawnienie pracy wydziałów komunikacji, spali na panewce.

Mamy tu więc sytuację użycia czasu przeszłego dla stwierdzenia, że coś z pewnością się nie uda. Jednakże kwerenda korpusowa ujawnia również takie przykłady, w których gramatyczna forma *futurum* jest użyta w syntaktycznym i pragmatycznym kontekście np. do wyrażenia wątpliwości, warunku, por.

(8) I diabli wiedzą, co robić, bo na gwałt muszę jeszcze skądś wytrzasnąć ze sto tysięcy. Inaczej wszystko spali na panewce. (NKJP; T. Dołęga-Mostowicz, *Drugie życie doktora Murka*)

(9) Teraz wszystko zależy od rodzaju zamków. Jeśli będą skomplikowane, pewnie cały plan spali na panewce. (NKJP; J. Głębski, *Kuracja*)

Same formy gramatyczne istnieją zatem w paradygmacie zwrotu, choć ich użycie ograniczone jest do niepodstawowej funkcji *futurum*. Tworząc słownik paradygmatów, uświadamialiśmy sobie niejednokrotnie, że skłonni jesteśmy, tak jak w wyżej przy-

wołanym opracowaniu, wykluczyć jakieś formy, bo nie znajdowaliśmy dla nich uzasadnienia w ciągach separowanych lub kongruentnych. Przykłady w rodzaju *spalić na panewce* ujawniają pewien dodatkowy problem przy tworzeniu słownika fleksyjnego: podaje on formy, które izolowane mogą budzić sprzeciw użytkownika, ale użyte w określonym kontekście są całkowicie zgodne z normą. Korpus takie właśnie konteksty niewątpliwie pomaga znaleźć. Podane przykłady potwierdzają zatem, z jednej strony, przydatność *corpus-based approach* w tworzeniu omawianego typu słownika, jego użyteczność w sprawdzaniu hipotez, z drugiej – pokazują konieczność weryfikacji uzyskanych danych na podstawie własnej kompetencji (analizy własności semantyczno-strukturalnych zwrotu). Takie refleksje na temat źródeł korpusowych w pracy nad opisywanym słownikiem nie są zaskakujące. Potwierdzają niewątpliwie stanowisko tych językoznawców, którzy twierdzą, że badacz, nie korpus, jest instancją rozstrzygającą, interpretatorem źródeł (zob. też uwagi Bańki (2001) na temat ćwiczenia z wyrażeniem *nie bez kozery*).

### 3. Podsumowanie

W artykule staraliśmy się opisać stopień przydatności różnych źródeł w tworzeniu słownika paradygmatów zwrotów frazeologicznych. Pierwsze z wymienionych na wstępie – metaźródła słownikowe – mają niewątpliwie zastosowanie jako istotny element w ustalaniu / weryfikowaniu listy jednostek. Drugie – źródła korpusowe – poza funkcją egzemplifikacyjną, były przez nas wykorzystywane w weryfikowaniu hipotez, choć w ograniczony sposób. Ograniczenia te (także przy preferowanym tu podejściu *corpus-based approach*) wynikają w sposób oczywisty z charakteru słownika. Bańko (2001: 43) w odniesieniu do słownika ogólnego zadał pytanie, czy w zakresie odmiany słowniki powinny odzwierciedlać system czy uzus. Odpowiedział na nie następująco: „słowniki mają odzwierciedlać system, ale nie powinny eksponować form nie poświadczonych w korpusie”. W przypadku słownika gramatycznego nie da się zastosować tej zasady, a więc różnicować statusu form, skoro paradygmaty w założeniu mają być podane wprost. Po drugie, jeśli przyjmujemy stanowisko systemowe, o włączeniu lub nie danej formy do paradygmatu rozstrzyga analiza właściwości znaczeniowo-strukturalnych zwrotu. Korpus natomiast może posłużyć do podania pewnych elementów charakterystyki pragmatycznej i znalezienia kontekstów nieprzewidzianych przez badacza (czy nawet z jego perspektywy zaskakujących). Jak zauważa Patrick Hanks (1990: 40), języki naturalne pełne są nieprzewidzianych faktów, które korpusy pomagają zidentyfikować. Jednocześnie stwierdzić trzeba, że korpusy lepiej niż w opisie gramatycznym sprawdzają się w badaniach sposobu użycia i frekwencji danej jednostki. Względem korpusu nie zajmujemy zatem żadnej z dwóch postaw skrajnych: tj. z jednej strony nie uznajemy introspekcji za jedyne wiarygodne narzędzie językoznawcy, a kwerendy korpusowej za „zbieranie motyli” nie posuwające wiedzy językowej badacza (jak twierdzi N. Chomsky), a z drugiej – nie kierujemy się absolutnym zaufaniem do korpusu i traktowaniem introspekcji jako najbardziej

ułomnego narzędzia lingwistycznego (jak twierdzą np. przedstawiciele kognitywizmu, por. np. Tummers, Heylen, Geeraerts 2005).

Na zakończenie można zadać sobie pytanie: czy sam korpus (w obecnej postaci) i źródła słownikowe wzbogacone o opracowania naukowe z dziedziny frazeologii są wystarczające, żeby stworzyć (opracować) słownik fleksyjny frazeologizmów? Odpowiedź musi brzmieć: nie. Z jednej strony wynika to z braku poświadczeń stanowiącej większości form w korpusach i milczenia słowników co do ich statusu. Z drugiej strony, o ile metaźródła w postaci słowników dostarczają nam „opracowanych” informacji na temat niektórych form fleksyjnych, o tyle sam korpus – tylko informacji o wystąpieniu form w tekstach. W żaden sposób nie rozstrzyga, czy było to użycie przewidywane przez system, czy innowacja językowa funkcjonalnie nieuzasadniona, czy wreszcie chociażby metajęzykowe użycie danego ciągu.

Podsumowując, w tworzeniu fleksyjnego słownika zwrotów większość decyzji co do statusu danej formy jest oparta właściwie wyłącznie na kompetencji językowej leksykografów, a nawet tam, gdzie źródła słownikowe czy opracowania naukowe wskazują status danej formy, każdorazowo powinien on być weryfikowany w oparciu o inne źródła (w tym korpusowe) i kompetencję językową słownikarzy.

## Bibliografia

- BAŃKO Mirosław (2001): *Z pogranicza leksykografii i językoznawstwa. Studia o słowniku jednojęzycznym*. – Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego.
- BAŃKO Mirosław (2012): NKJP w oczach leksykografa. – [w:] Adam PRZEPIÓRKOWSKI, Mirosław BAŃKO, Rafał L. GÓRSKI, Barbara LEWANDOWSKA-TOMASZCZYK (red.): *Narodowy Korpus Języka Polskiego*. – Warszawa: PWN, 283–290.
- BĄBA Stanisław (1998): Uwagi o odmianie polskich zwrotów przez czasy. – *Poznańskie Spotkania Językoznawcze III*, 7–14.
- BIBER Douglas (2015): Corpus-based and corpus-driven analyses of language variation and use. – [w:] Bernd HEINE, Heiko NARROG (red.): *The Oxford Handbook of Linguistic Analysis*. – Oxford, New York: Oxford University Press, 159–191.
- BOGUSŁAWSKI Andrzej (1976): O zasadach rejestracji jednostek języka. – *Poradnik Językowy* 8, 356–364.
- BOGUSŁAWSKI Andrzej, GARNYSZ-KOZŁOWSKA Teresa (1979): *Addenda do frazeologii polskiej. Zeszyt wstępny*. – Edmonton: Carbondale.
- BOGUSŁAWSKI Andrzej, WAWRZYŃCZYK Jan (1993): *Polszczyzna jaką znamy. Nowa sonda słownikowa*. – Warszawa: Uniwersytet Warszawski.
- BOGUSŁAWSKI Andrzej, DANIELEWICZOWA Magdalena (2005): *Verba polona abscondita. Sonda słownikowa III*. – Warszawa: Elma Books.
- CONRAD Susan (2010): What can a corpus tell us about grammar? – [w:] Anne O'KEEFFE, Michael MCCARTHY (red.): *The Routledge Handbook of Corpus Linguistics*. – London–New York, NY: Routledge, 227–240.
- CZEREPOWICKA Monika (2014): SEJF – Słownik elektroniczny jednostek frazeologicznych. – *Język Polski* XCIV, 1, 116–129.
- CZEREPOWICKA Monika, KOSEK Iwona, PRZYBYSZEWski Sebastian (2014): O projekcie elektronicznego słownika odmiany frazeologizmów czasownikowych. – *Polonica* XXXIV, 115–123.

- ČERMÁK František (2003): Source materials for dictionaries. – [w:] Piet VAN STERKENBURG (red.): *A practical guide to lexicography*. – Amsterdam–Philadelphia: John Benjamins Publishing, 18–25.
- GÓRSKI Rafał L. (2012): Zastosowanie korpusów w badaniu gramatyki. – [w:] Adam PRZEPIÓRKOWSKI, Mirosław BAŃKO, Rafał L. GÓRSKI, Barbara LEWANDOWSKA-TOMASZCZYK (red.): *Narodowy Korpus Języka Polskiego*. – Warszawa: PWN, 291–299.
- GRABOWSKI Łukasz (2015): O frazeologii z perspektywy językoznawstwa korpusowego. Przegląd głównych nurtów badawczych ostatniego dwudziestolecia w Wielkiej Brytanii i USA. – *Problemy frazeologii europejskiej* X, Łask: Oficyna Wydawnicza LEKSEM, 23–48, online: <http://pfe.leksem.pl/wp-content/uploads/Grabowski-1.pdf>.
- HANKS Patrick (1990): Evidence and intuition in lexicography. – [w:] Jerzy TOMASZCZYK, Barbara LEWANDOWSKA-TOMASZCZYK (red.): *Meaning and Lexicography*. – Amsterdam–Philadelphia: John Benjamins Publishing Company, 31–41.
- KILGARIFF Adam (2013): Using corpora as data source for dictionaries. – [w:] Howard JACKSON (red.): *The Bloomsbury companion to lexicography*. – London–New York: Bloomsbury Academic, 77–96.
- KOSEK Iwona (2013): Paradygmaty zwrotów frazeologicznych – problemy opisu leksykograficznego. – [w:] Gabriela DZIAMSKA-LENART, Jarosław LIBEREK (red.): *Perspektywy współczesnej frazeologii polskiej. Między teorią a praktyką frazeologiczną*. – Poznań: Wydawnictwo Naukowe UAM, 51–61.
- LEWICKI Andrzej Maria (1976): *Wprowadzenie do frazeologii syntaktycznej. Teoria zwrotu frazeologicznego*. – Katowice: Uniwersytet Śląski.
- LEWICKI Andrzej Maria (1986): Składnia związków frazeologicznych. – *Biuletyn Polskiego Towarzystwa Językoznawczego* XL, 75–82.
- LEWICKI Andrzej Maria, PAJDZIŃSKA Anna (2001): Frazeologia – [w:] Jerzy BARTMIŃSKI (red.): *Współczesny język polski*. – Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 315–333.
- LEWICKI Andrzej Maria, PAJDZIŃSKA Anna, REJAKOWA Bożena (1987): *Z zagadnień frazeologii. Problemy leksykograficzne*. – Warszawa: PWN.
- LIBEREK Jarosław (2006): Zróżnicowanie aspektowe komponentów werbalnych w stałych związkach frazeologicznych. Uwagi wstępne. – [w:] Stanisław BĄBA, Piotr FLICIŃSKI, Jarosław LIBEREK (red.): *Z zagadnień frazeologii, stylistyki i kultury języka*. – Poznań: Wydawnictwo Poznańskie.
- LUBAŚ Władysław, red. (1988): *Wokół słownika współczesnego języka polskiego. Materiały konferencji w Paszówce 26–28 XI 1986 r.* – Wrocław: Ossolineum, PAN.
- MARCINIAK Małgorzata, RABIEGA-WIŚNIEWSKA Joanna, SAVARY Agata, WOLIŃSKI Marcin, HELIASZ Celina (2009): Constructing an Electronic Dictionary of Polish Urban Proper Names. – [w:] *Recent Advances in Intelligent Information Systems (Proceedings of the Balto-Slavonic Natural Language Processing Workshop, Kraków)*. – Warszawa: Oficyna Wydawnicza EXIT, 743–749.
- MARKOWSKI Andrzej (2005): *Kultura języka polskiego. Zagadnienia leksykalne*. – Warszawa: PWN.
- MOON Rosamund (1998): *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. – Oxford: Oxford University Press.
- PRZYBYSZEWSKI Sebastian (2015): Some problems with the description of paradigms of Polish verbal multi-word units. – [w:] Enrique GUTIÉRREZ RUBIO, Marta FALKOWSKA, Ekaterina KISLOVA, Marzena STĘPIEŃ (red.): *Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV)*, t. 18, Wiesbaden: Harrassowitz Verlag, 213–223.
- SALONI Zygmunt (2001): *Czasownik polski. Odmiana, słownik*. – Warszawa: PWN.
- SINCLAIR John (1995): *Corpus, concordance, collocation*. – Oxford: Oxford University Press.
- TUMMERS Jose, HEYLEN Kris, GEERAERTS Dirk (2005): Usage-based approaches in Cognitive Linguistics: A technical state of the art. – *Corpus Linguistics and Linguistic Theory* 1 (2), 225–261.
- WALTER Elizabeth (2010): Using corpora to write dictionaries. – [w:] Anne O'KEEFE i Michael MCCARTHY (red.): *The Routledge Handbook of Corpus Linguistics*. – London–New York, NY: Routledge, 428–443.
- ŻMIGRODZKI Piotr, opr. (2015): *Wielki słownik języka polskiego PAN. Zasady opracowania*, wyd. II. – Kraków.

**Słowniki i korpusy**

- ISJP – Mirosław BAŃKO, red. (2000): *Inny słownik języka polskiego*, I–II. – Warszawa: PWN.
- NKJP – Adam PRZEPIÓRKOWSKI, Mirosław BAŃKO, Rafał L. GÓRSKI, Barbara LEWANDOWSKA-TOMASZCZYK (red.): *Narodowy Korpus Języka Polskiego*. – Warszawa: PWN; <http://www.nkjp.pl>.
- PSF – Stanisław BĄBA, Jarosław LIBEREK, Gabriela DZIAMSKA-LENART, red. (2001): *Podręczny słownik frazeologiczny języka polskiego*. – Warszawa: PWN.
- SFJP – Stanisław SKORUPKA (1985): *Słownik frazeologiczny języka polskiego*. – Warszawa: Wiedza Powszechna, wydanie IV.
- SFWP – Stanisław BĄBA, Jarosław LIBEREK, red. (2001): *Słownik frazeologiczny współczesnej polszczyzny*, Warszawa: PWN.
- SGJP – Zygmunt SALONI, Marcin WOLIŃSKI, Robert WOŁOZ, Włodzimierz GRUSZCZYŃSKI, Danuta SKOWROŃSKA, red. (2015): *Słownik gramatyczny języka polskiego*. Wyd. III on-line, <http://sgjp.pl>.
- SJPD – Witold DOROSZEWSKI, red. (1958-1969): *Słownik języka polskiego*, I–XI. – Warszawa: PWN.
- SJPSz – Mieczysław SZYMCZAK, red. (1996): *Słownik języka polskiego*, I–III. – Warszawa: PWN.
- SWJP – Bogusław DUNAJ, red. (1996): *Słownik współczesnego języka polskiego*. – Warszawa: Wydawnictwo Wilga.
- USJP – Stanisław DUBISZ, red. (2003): *Uniwersalny słownik języka polskiego*, I–IV. – Warszawa: PWN.
- WSF – Anna KŁOSIŃSKA, Elżbieta SOBOL, Anna STANKIEWICZ, red. (2005): *Wielki słownik frazeologiczny z przysłowiami*. – Warszawa: PWN.
- WSFJP – Piotr MÜLDNER-NIECKOWSKI (2003): *Wielki słownik frazeologiczny języka polskiego*. – Warszawa: Świat Książki.
- WSJP – Piotr ŻMIGRODZKI, red. nauk.: *Wielki słownik języka polskiego*, <http://www.wsjp.pl>