

Użyteczność nowych zasobów cyfrowych w analizie diachronicznej ekspresywnych jednostek leksykalnych

JAKUB BOBROWSKI*

1. Wprowadzenie

Leksyka ekspresywna od dawna przyciągała uwagę językoznawców, w tym również badaczy polszczyzny. Wśród prac polskich autorów poświęconych słownictwu nacechowanemu emocjonalnie znaleźć można zarówno pozycje zorientowane teoretycznie, z klasyczną monografią Stanisława Grabiasa (1981) na czele, jak i liczne opracowania materiałowe, poświęcone językowi współczesnemu (np. Burkacka 2001; Buttler 1979; Kreja 1969), a także historycznemu (np. Książek-Bryłowa 2001; Walczak 1992)¹. Wśród tych ostatnich szczególnie miejsce zajmuje książka Artura Rejtera (2006), przynosząca szerokie omówienie formalnej, semantycznej i pragmatycznej ewolucji

* <https://orcid.org/0000-0002-1096-4677>, Instytut Języka Polskiego Polskiej Akademii Nauk, Kraków, Polska, jakub.bobrowski@ijp.pan.pl

¹ Podaję tu oczywiście tylko przykładowe opracowania. Wyczerpujące zestawienie bibliograficzne znaleźć można w monografii A. Rejtera (2006).

polskiej leksyki ekspresywnej na przestrzeni kilkuset lat z uwzględnieniem tła kulturowego. Choć o przeszłości i współczesności wskazanej klasy słownictwa wiemy już bardzo wiele, warto rozważyć, w jakim stopniu rozwój nowych nurtów metodologicznych w językoznawstwie może się przyczynić do wzbogacenia dotychczasowych ustaleń. Szczególnie ciekawe perspektywy zdaje się otwierać wykorzystanie w badaniach nad polszczyzną technologii informatycznych. Mam tu na myśli zarówno wąsko rozumianą lingwistykę komputerową, jak i tworzenie zdigitalizowanych baz danych w ramach tzw. nowej humanistyki. W niniejszym artykule chciałbym pokazać, że powstające współcześnie cyfrowe zasoby językowe, odpowiednio zaprogramowane, stanowią cenne źródło nowych danych do badań diachronicznych nad leksemami ekspresywnymi. Powinny się one przyczynić do uściślenia i dopełnienia wielu wcześniej sformułowanych i powszechnie przyjmowanych sądów.

Lingwistyka komputerowa wypracowała szereg narzędzi i zasobów umożliwiających automatyczne przetwarzanie i analizowanie języka². Centralne miejsce zajmują wśród nich korpusy tekstów. Dla języka polskiego dysponujemy już kilkoma korpusami reprezentującymi jego poszczególne fazy rozwojowe. Polszczyznę w szerokim rozumieniu tego słowa współczesną (XX–XXI w.) znajdziemy w Narodowym Korpusie Języka Polskiego (NKJP, por. Przepiórkowski i in. 2012). Teksty pochodzące z wcześniejszych okresów stały się podstawą: Korpusu polszczyzny XVI wieku (KP XVI), Elektronicznego korpusu tekstów polskich XVII i XVIII w. (Korba) i Korpusu języka polskiego lat 1830–1918 (KJP XIX). Zgłoszono również projekt opracowania Narodowego Korpusu Diachronicznego Polszczyzny, który połączy różne zasoby diachroniczne (por. Król i in. 2019)³. W tej chwili jedynie NKJP można uznać za dzieło ukończone i w pełni reprezentatywne (choć w coraz mniejszym stopniu dla języka najnowszego). Pozostałe przedsięwzięcia znajdują się w fazie realizacji bądź mają postać załączkową. Korpusy bez wątpienia stanowią cenną pomoc w analizie ewolucji słownictwa, również ekspresywnego, nie tylko bowiem gromadzą ogromną liczbę tekstów, ale oferują ponadto zaawansowany język zapytań, dzięki czemu daje się je przeszukiwać

² Dotychczasowe osiągnięcia polskiej lingwistyki komputerowej oraz perspektywy jej rozwoju przedstawili Magdalena Derwojedowa (2017) i Maciej Ogrodniczuk (2017).

³ W sieci dostępnych jest jeszcze kilka mniejszych zasobów diachronicznych polszczyzny, nie mają one jednak istotnego znaczenia dla problematyki podjętej w niniejszym artykule. Ich omówienie znaleźć można w opracowaniu Patrycji Pałki i Agaty Kwaśnickiej-Janowicz (2017: 162–168).

dużo szybciej i efektywniej niż „surowe” zbiory tekstów, nawet te dostępne w postaci zdigitalizowanej. Wydaje się jednak, że użyteczność źródeł tego rodzaju najbardziej uwidacznia się w badaniach zorientowanych gramatycznie, w poszukiwaniach leksykologicznych pozostaje natomiast ograniczona. Wynika to z faktu, że celem tworzenia korpusu ogólnego jest uchwycenie ogólnych tendencji obecnych w systemie językowym danej epoki, a nie idiosynkrazji. Aby to osiągnąć, należy go zrównoważyć pod względem stylistycznym, gatunkowym, geograficznym itp. Wskutek selekcji materiału zjawiska rzadsze, których występowanie zawężone jest do pewnych odmian języka, takie jak nacechowane ekspresywnie jednostki leksykalne, mogą być reprezentowane w niewystarczającym stopniu. Choć zatem w badaniach nad życiem wyrazów korpusy mogą okazać się pomocne, nie należy ich roli absolutyzować i warto sięgać również po inne cyfrowe zasoby językowe.

Dla dociekań leksykologicznych szczególną wartość reprezentują rozwijane coraz intensywniej biblioteki cyfrowe. Powstają one w związku z ogólnym zwrotem cyfrowym w kulturze i naukach humanistycznych, w przeciwieństwie zatem do zasobów tworzonych w ramach lingwistyki komputerowej nie są dostosowane specjalnie do potrzeb językoznawcy. Jednak również badacz języka jest w stanie wydobyć z nich cenne informacje. Zdigitalizowanych zasobów bibliotecznych nie można wprawdzie przeszukiwać za pomocą tak wyrafinowanych narzędzi, jak ma to miejsce w przypadku korpusów, nie oznacza to jednak, że ekscerpujący je lingwista skazany jest na pracę zupełnie mechaniczną. Najlepsze biblioteki cyfrowe oferują już dzisiaj wyszukiwarki technicznie całkiem zaawansowane, które umożliwiają pozyskanie materiału przefiltrowanego za pomocą różnorodnych kryteriów: chronologicznych, gatunkowych, autorskich, tematycznych. Obecnie z punktu widzenia badacza słownictwa najbardziej obiecującymi przedsięwzięciami wydają się Polona, czyli Cyfrowa Biblioteka Narodowa (www.polona.pl), oraz usługa Google Books (books.google.pl). Polona to polska biblioteka cyfrowa gromadząca zdigitalizowane obiekty ze zbiorów Biblioteki Narodowej, Biblioteki Jagiellońskiej oraz szeregu innych polskich instytucji kulturalnych i naukowych. Z kolei Google Books jest usługą oferowaną w ramach serwisu Google, która umożliwia dostęp do książek i czasopism zeskanowanych i przekonwertowanych na wersję tekstową za pomocą oprogramowania OCR, a pozyskanych ze zbiorów partnerskich placówek bibliotecznych z całego świata. Liczba dokumentów zgromadzonych w ramach wymienionych przedsięwzięć wielo-

krotnie przekracza to, co badaczowi leksyki oferują korpusy. Moje dotychczasowe doświadczenia eksploracyjne pokazują, że Polona, licząca obecnie (w 2020 r.) ponad 3 miliony obiektów (taka liczba podana jest na stronie projektu), przedstawia szczególną wartość jako reprezentacja polszczyzny od połowy XVIII do połowy XX w. O ile wcześniejsze fazy rozwojowe języka polskiego, ze względu na ograniczony ilościowo i w miarę zamknięty kanon dostępnych źródeł, zostały lub zostaną udokumentowane w satysfakcjonujący sposób w wymienionych wcześniej korpusach oraz słownikach historycznych (SS^{Stp}, SPXVI, ESXVII), o tyle słownictwo wskazanego okresu okazuje się zdumiewająco słabo poznane. Dość wspomnieć, że bardzo wiele wyrazów nienotowanych w SJP^{Dor} jest poświadczonych w zasobach Polony⁴. Z kolei Google Books oferuje aktualnie dostęp do ponad 40 milionów tytułów w ponad 400 językach (Lee 2019). Wprawdzie publikacje polskojęzyczne stanowią tylko niewielką część tego zasobu i gromadzone są w sposób raczej przypadkowy, ale z punktu widzenia badacza polszczyzny i tak stanowią ogromną kolekcję danych, szczególnie cenną jako reprezentacja leksyki dwudziestowiecznej. Na podstawie własnych doświadczeń badawczych mogę powiedzieć, że dla tego zakresu czasowego Google Books bywa źródłem bardziej użytecznym zarówno od Polony, jak i NKJP.

Dość bogatym i wartym przeglądania zasobem okazują się również Wikiźródła (pl.wikisource.org), kolekcja tekstów tworzona w ramach Wikipedii. Wprawdzie, podobnie jak ma to miejsce w Google Books, ich dobór wydaje się dosyć przypadkowy, a ponadto duża część z nich nie jest wciąż zoceerowana i została udostępniona wyłącznie w postaci skanów, niemniej jednak językoznawca może napotkać w tym zbiorze pozycje cenne informacyjnie, nieobecne w innych bazach danych. Pozostałe biblioteki cyfrowe (ich omówienie znaleźć można w pracy Pałka, Kwaśnicka-Janowicz 2017: 33–37) na obecnym etapie zdają się już mieć dla badań leksykologicznych dużo mniejsze znaczenie.

Twórcy bibliotek cyfrowych starają się gromadzić wszystkie dostępne dokumenty, nie przeprowadzając selekcji, której celem byłoby zrównoważenie zasobu pod względem stylistycznym, geograficznym, gatunkowym itp. Równocześnie dostępne teksty pozostają ogromnie zróżnicowane, zarówno językowo, jak i pod względem nośnika (książki, prasa, druki ulotne, rękopisy)

⁴ Na istotne ograniczenia materiałowe SJP^{Dor} od dawna zwracał uwagę Jan Wawrzyńczyk (por. np. Wawrzyńczyk 2010).

czy też jako dokumentacja życia społecznego. Dzięki temu istnieje większa szansa na odszukanie poświadczeń wyrazów rzadszych, w tym ekspresywnych, niż ma to miejsce przy korzystaniu z korpusów. Niestety omawiane zasoby źródłowe nie są pozbawione wad. Przede wszystkim wyszukiwarki bibliotek cyfrowych bywają zawodne, jeśli chodzi o rozpoznawanie form fleksyjnych leksemów polskich. Zdarza się też dość często, że błędnie interpretują pewne zapisy (zwłaszcza niedoskonale technicznie) jako poświadczenia danego leksemu. Wreszcie w dość ograniczonym stopniu radzą sobie z wyszukiwaniem jednostek nieciągłych. Wszystko to powoduje, że badacz, jeśli zależy mu na pozyskaniu danych naprawdę rzetelnych, zmuszony jest niejednokrotnie do dosyć żmudnego, ręcznego weryfikowania wyników. Nie można też zapominać o tym, że zasoby bibliotek cyfrowych, w tym Polony i Google Books, są nieustannie rozbudowywane, wzbogacane o nowe pozycje, co niesie ze sobą ryzyko szybkiej falsyfikacji poczynionych ustaleń. W przypadku drugiej z wymienionych platform dodatkowym ograniczeniem jest brak dostępu do pełnej wersji niektórych zdigitalizowanych tekstów ze względu na prawa autorskie.

Biorąc pod uwagę wymienione zalety i wady różnego rodzaju zasobów cyfrowych, należy stwierdzić, że niewskazane jest ograniczanie się w poszukiwaniach leksykologicznych do jednego ich typu czy tym bardziej jednego, konkretnego źródła, każda z baz bowiem może dostarczyć danych, które nie występują w pozostałych. W sposób szczególny zastrzeżenie to dotyczy badań nad słownictwem rzadszym, słabiej reprezentowanym w tekstach pisanych, a do tej grupy możemy zaliczyć ekspresywizmy.

W dalszej części artykułu chciałbym pokazać, jak cyfrowe bazy danych można wykorzystać w diachronicznej analizie funkcjonowania w tekstach konkretnych, nacechowanych emocjonalnie jednostek leksykalnych. Uważam, że szczególnie warto do nich sięgać w ramach następujących procedur badawczych:

1. Chronologizacji ekspresywnych jednostek leksykalnych (ustalenie, od kiedy istnieją w polszczyźnie).
2. Rekonstrukcji procesu upowszechniania się w polszczyźnie ekspresywnych jednostek zapożyczonych.
3. Rekonstrukcji procesu upowszechniania się w polszczyźnie nowych znaczeń ekspresywnych wyrazów już istniejących.

4. Falsyfikacji utrwalonych w leksykografii i opracowaniach naukowych kwalifikacji stylistycznych jednostek ekspresywnych.

2. Chronologizacja leksyki ekspresywnej

Najprostszy sposób wykorzystania zasobów cyfrowych w badaniu historii słownictwa ekspresywnego polega na cofnięciu chronologizacji jednostek, co do których można by domniemywać, że są tworamii stosunkowo młodymi. Poszukiwania lingwochronologiczne wspomagane komputerowo nie są oczywiście w polskim językoznawstwie czymś nowym. Zostały zainicjowane już ponad dziesięć lat temu przez Piotra Wierzchońia i Jana Wawrzyńczyka (por. zwłaszcza Wierzchoń 2008). Za najważniejsze dokonanie badaczy należy uznać Narodowy Fotokorpus Języka Polskiego (NFJP), w którym zgromadzono wielką liczbę fotocytatów, czyli udokumentowanych fotograficznie poświadczeń użycia wyrazów i ich połączeń, pochodzących z lat 1900–2000. Twórcom Fotokorpusu udało się odkryć wiele jednostek leksykalnych nieopisanych dotąd leksykograficznie, zebrana przez nich dokumentacja pozwoliła też sfalsyfikować szereg hipotez datacyjnych formułowanych wcześniej w odniesieniu do słownictwa dwudziestowiecznej polszczyzny. Było to bez wątpienia istotne osiągnięcie, ale rozwój nowych przedsięwzięć digitalizacyjnych sprawił, że obecnie dysponujemy również innymi zasobami dostarczającymi cennych informacji na temat rozwoju polskiej leksyki w minionym stuleciu. Dotyczy to oczywiście także jednostek ekspresywnych. Pokażemy to na przykładzie dwóch słów funkcjonujących w świadomości społecznej jako symbole wulgaryzacji języka w ostatnich latach – przymiotnika *zajebisty* oraz przysłowka *zajebiście*. W znanym cyklu opracowań *Nowe słownictwo polskie*, przygotowywanym przez zespół Teresy Smółkowej i dokumentującym najświeższą warstwę leksyki na podstawie ręcznej analizy prasy, wskazane wyrazy pojawiły się dopiero w tomach za lata 1993–2000 (por. NSP 2006). Cytat ilustrujący hasło *zajebisty* pochodzi z 1998 r., a *zajebiście* – z 1993 r. Analiza materiałów udostępnianych w ramach usługi Google Books pozwoliła znaleźć starsze poświadczenia. Po raz pierwszy oba słowa użyte zostały w listach poety Ryszarda Milczewskiego-Bruna z 1970 r., wydanych pośmiertnie w 1989 r.:

- (1) To tyle. Napisz rychło. Póki jestem w formie. (A mam sporo nowych, być może zajebistych – wierszy).

Pozdrawiam serdecznie
Bruno

(Ryszard Milczewski-Bruno, *Poezja. Proza. Listy*, 1989, books.google.pl)⁵

- (2) Szkoda, że Ciebie nie ma. Siedzę. Piję „Gryf-vodkę”! Śpiewa – zajebicie – Okudźawa (a raczej: jego).

(Ryszard Milczewski-Bruno, *Poezja. Proza. Listy*, 1989, books.google.pl)

Można napotkać jednak również inne wystąpienia wcześniejsze niż dotychczas proponowane datacje. Co ciekawe, pochodzą one z prozy fabularnej, por.:

- (3) Mundur, bo mundur – kombinuje – ale co, zapięty na kulki? Bo jakby co, to ten zajebisty szmondak mi nie pożałuje.

(Maciej Zenon Bordowicz, *Orkiestranty*, 1974, books.google.pl)

- (4) [...] budziłem się oblany, lub okakany, nie śmiej się, chuju, cierpiałem dla forsy, dla takiej prawdziwej, ogromnej, zajebicie amerykańskiej, nieraz w supermarkecie obserwowałem kobiety, wkładające do wózków te idiotyczne galaretki i nie mogłem się nadziwić, pakiet kosztował dwadzieścia centów, skąd te miliony?

(Andrzej Brycht, *Opowieści z tranzytu*, 1986, books.google.pl)

Obecność wskazanych jednostek w prywatnych listach można uznać za potwierdzenie samego ich istnienia w polszczyźnie potocznej (i to raczej w jej niższym rejestrze) przynajmniej od końca lat 60. Z kolei użycie omawianych słów w tekstach literackich jako wykładników kolokwializacji sugeruje, że w latach 70. i 80. były już prawdopodobnie mocno rozpowszechnione.

3. Upowszechnianie się ekspresywnych jednostek zapożyczonych

Wspomagane komputerowo poszukiwania najstarszych poświadczeń jednostek ekspresywnych mogą również rzucić nowe światło na proces przenikania do polszczyzny wyrazów nacechowanych mających swe źródła w językach obcych. Za przykład posłuży tu rzeczownik *szantrapa*. Wyraz ten, mający oczywisty źródłosłów rosyjski (*шантрапа*), w słownikach współczesnej pol-

⁵ We wszystkich przykładach zachowana została pisownia oryginalna. Podkreślenia pochodzą od autora.

szczyzny definiowany jest jako ‘kobieta kłótliva i niesympatyczna’ (SJPDun, USJP), choć jeszcze w SJPDor opisany został jako obelżywe określenie ‘kobiety niechlujnej, niemoralnej’. Wskazuje to na rozchwianie semantyczne, co jednak nie dziwi, zważywszy, że słowo używane jest w języku potocznym, w dodatku przede wszystkim jako wyzwisko. Zarówno definicje słownikowe, jak i materiał zawarty w NKJP pokazują, że w języku polskim leksem *szantrapa* odnoszony jest wyłącznie do kobiet, inaczej niż ma to miejsce w języku rosyjskim, w którym – jak pisze Ludwika Jochym-Kuszlukowa (2002: 104) – może być określeniem mężczyzny, kobiety i zbiorowości, inna jest także jego rosyjska semantyka (‘drań, hołota’). Omawiany wyraz wiązany bywa z powojenną falą rusycyzmów (por. wpis pt. *Skąd się wzięła szantrapa?* na stronie jednej z internetowych poradni językowych: <http://www.poradnia-jezykowa.uz.zgora.pl/wordpress/?p=768>), częściej jednak badacze wskazują na pośrednictwo gwar miejskich (zwłaszcza warszawskiej) i przestępczych, co sugeruje oczywiście wcześniejszą chronologię (por. Jochym-Kuszlukowa 2002; Kania 1995; Milejkowska 1966).

Analiza danych cyfrowych pozwoliła ustalić, że w tekstach pisanych *szantrapa* zaczęła być używana w latach tuż przed I wojną światową:

- (5) – Wiesz, Antoś, jest Mańka Fijoł z Leosiem.
– Gdzie?
– Spójrzyj, za tą szantrapą w szynszylach.
– Długo będziesz się jeszcze kręciła?
– Ona mnie obraziła.

(*Na dzień dobry*, „Kurjer Poranny”, nr 316, 1911, polona.pl)

- (6) Antek klawisz, chłop nie gapa.
(Choć ma we łbie też i śmieci).
Nabujała go szantrapa,
Która męża ma i dzieci,
Że ma forsy taką miarę,
Że starczy na kamienicę...

(*Za granicę. Kuplety*, „Trubadur Polski”, nr 50, 1912, polona.pl)

Można zatem domniemywać, że omawiany leksem przeniknął do polszczyzny pod koniec XIX lub na początku XX w., przy czym zakres jego użycia musiał być wówczas jeszcze dosyć ograniczony, nie został bowiem odnotowany w szóstym tomie tzw. słownika warszawskiego (SW), obejmującym literę S, wydanym w 1915 r. Przywołane użycia pochodzą z prasy

warszawskiej, z tekstów satyrycznych wyraźnie stylizowanych na „mowę ulicy”, co potwierdzałoby formułowany wcześniej w pracach językoznawczych pogląd, że z języka rosyjskiego *szantrapa* przeniknęła najwcześniej do gwary miejskiej Warszawy. Trzeba jednak zaznaczyć, że już w dwudziestoleciu międzywojennym wyraz pojawia się w źródłach prasowych pochodzących z innych rejonów Polski, np. Częstochowy:

- (7) Ty *szantrapo*, urwipelciu jeden, albo franki sobie kup, albo zaraz światło gaś, bo tu ucziwe ludzie mieszkają [...].

(*Obrazki sądowe*, „Słowo Częstochowskie”, nr 142, 1934, polona.pl)

Katowic:

- (8) [...] kaźden jeden zakochany ślepa komenda i ofiara losu bita i kopana się robi i najgorszą cholerę *szantrapę* można takiemu wtranżolić [...].

(*Dzieje występnej miłości*, „Nowy Czas”, nr 227, 1934, polona.pl)

Nowogródka:

- (9) To on mnie do tego namówił! Bandyta, złodziej! a wszystko dla tej *szantrapy*, dla tej złodziejki!

(Felicja Romanowska, *Wileńska powieść kryminalna*, „Kurjer Nowogródzki”, nr 79, 1933, polona.pl)

Poznania:

- (10) [...] aliści niech cygańskie romanse zagrają lub co nie daj Boże, zaśpiewa urocza, może nie tyle „chanteuse”, co *szantrapa* to brawa braciszku takie, że ogłuchnąć można z kretesem.

(*Kaunas czyli mój pierwszy zajazd na Litwę*, „Kurjer Poznański”, nr 191, 1925, polona.pl)

Wcześniej trafił również do literatury popularnej pozbawionej wyraźnie regionalnego zabarwienia, najstarsze użycie tego typu, które udało mi się znaleźć, pochodzi bowiem z roku 1924:

- (11) Co sobie o mnie ta dzierlatka pomyśli, jeżeli jej się uda rzeczywistą stronę medalu ukryć przed Kręcickim?... Jakie będzie miała wyobrażenie o suwerennym przedstawicielu sejmu, który na rozkaz pierwszej z brzegu *szantrapy* pakuje się pod łóżko, jak jaki kundys, albo wyżeł tresowany???

(L. J. Kazimierski, *Nasi suwereni. O posłach, sejmie, oraz ministrach historia bez mała prawdziwa*, 1924, polona.pl)

Pokazuje to, że analizowany leksem bardzo szybko stał się ekspresywi-
zmem o ponadregionalnym zasięgu.

W najwcześniejszym okresie funkcjonowania słowa w polszczyźnie domi-
nują już całkowicie, podobnie jak ma to miejsce dzisiaj, użycia w stosunku
do kobiet. Odniesienia męskie i kolektywne są nieco późniejsze, zdarzają
się rzadko i, co ciekawe, pełnią przede wszystkim funkcję ewokantów ro-
syjskości w tłumaczeniach literatury rosyjskiej bądź tekstach w jakiś sposób
dotyczących Rosji, np.:

- (12) Lebedincew gniewnie pociągnął fajkę i na chwilę zamyślił się:
– No, a potem jasne... Powyłażała wszelaka s z a n t r a p a ... Zaczęli pełzać, jak robaki
na trupie... Lwow, Milukow, błazen grochowy Kierenskiej [...].

(Ūrij Galič, *Chińskie cienie*, 1928, polona.pl)

- (13) Okaziciel niniejszego jest rzeczywiście okazicielem, a nie jakąś tam s z a n t r a p ą .

(Michaił Bułhakow, *Fatalne jaja*, 1929, polona.pl)

Ponadto pojawiają się czasem w prasie ze wschodnich terenów Rzeczy-
pospolitej jako wykładniki stylizacji na polszczyznę kresową, pozostającą pod
oczywistym wpływem wschodniosłowiańskim, por.:

- (14) Salonowa wychowawnia nam pokazuje! Co, znacysia my chamy ostatnie i nie
znamisia na delikacjach, tylko on, – s z a n t r a p a nieszczensny!

(Wincuk Markotny, *Na ulicy Legionowej*, „Słowo”,
nr 353, 1936, polona.pl)

Taki rozkład materiału sugeruje, że polska *szantrapa* jako obraźliwe
określenie kobiety, choć genetycznie rosyjska, bardzo szybko, bo już przed
II wojną światową, oddzieliła się od swoich korzeni i upowszechniła jako
ogólnopolski kolokwializm.

4. Upowszechnianie się ekspresywnych neosemantyzmów

Dzięki wykorzystaniu zasobów cyfrowych można również z wyjątkową
precyzją rekonstruować proces wykształcania się nowych znaczeń ekspresyw-
nych wyrazów. Przykładem może być silnie nacechowana jednostka *ćpać*.
Czasownik ten funkcjonuje w polszczyźnie od XVI w. (SPXVI podaje przykład
użycia z 1588 r.), jednak przez wieki używano go wyłącznie w znaczeniu
‘jeść łapczywie’. Tymczasem współcześnie odnoszony jest przede wszyst-

kim do zażywania narkotyków. Tego rodzaju użycia dominują zdecydowanie w NKJP. Przykłady dotyczące jedzenia, które można znaleźć w korpusie, pochodzą głównie ze starszych tekstów literackich (np. *Jeziro Bodeńskie* Stanisława Dygata, *Pokolenie* Bogdana Czeszki). Z kolei bazujący na NKJP *Wielki słownik języka polskiego PAN* (WSJP, por. Żmigrodzki i in. (red.) 2018), najnowszy z dużych słowników ogólnych, umieścił znaczenie ‘zażywać narkotyki’ na pierwszym miejscu, podczas gdy ISJP i USJP dawały jeszcze pierwszeństwo ‘łapczywemu jedzeniu’. Powyższe fakty skłaniają do postawienia pytania, od kiedy czasownika *ćpać* używa się w odniesieniu do przyjmowania środków odurzających. W Korpusie najstarsze przykłady pochodzą dopiero z 1992 r., z tekstu nadesłanego na konkurs „Temat Otwarty”, np.:

- (15) Poddawał się wszelkim rygorom monarowskiej „kuracji”, przy czym chodził do szkoły, przy czym smakował na nowo życie. Nie ćpał, nie pił, nie palił.

(NKJP: Małgorzata Najgrodzka, *Szukanie siebie*, „Gazeta Wyborcza”, 03.08.1992)

Analiza danych zwartych w Polonie oraz Google Books wykazała, że aż do początku lat 70. czasownik *ćpać* używany był wyłącznie w znaczeniu tradycyjnym, por. następujące przykłady z roku 1974:

- (16) [...] ćpali już ten klopsik ci, co go zdobyli. Zobaczywszy, co ćpają, czym ćpają i na czym, wyszedłem z dworca, tym bardziej że kolejka nie posunęła się ani o krok.

(„Twórczość”, t. 30, 1974, books.google.pl)

- (17) [...] nikt obżarstwem wstrętnym i nienasyconym nie zgrzeszył, flaków nie napełnił najulubieńszymi kęsami, nie ćpał i nie mlaskał ochotnie przy najlepszych dorodnych, wonnych gorących potrawach [...].

(„Miesięcznik Literacki”, t. 9, 1974, books.google.pl)

Najstarszy cytat odnoszący się do narkotyków pochodzi z roku 1975:

- (18) A w Nowym Jorku ogromnie dużo naszych nadużywało narkotyków. Przejechaliśmy przez kraj samochodem. [...] Gdy dotarliśmy do San Francisco, zobaczyliśmy jakby nowy świat. Ulice wyglądały jak w czasie karnawału, pełne nieprawdopodobnie młodych ludzi, którzy ćpali, wyczuwali się i za wszelką cenę rwali się do życia.

(„Dialog”, t. 20, 1975, books.google.pl)

Co ciekawe, pochodzi on z tłumaczenia i odnosi się do obcych realiów. Ma też charakter w znacznym stopniu wyjątkowy, kolejne bowiem poświadczenia interesującego nas znaczenia pochodzą z lat 80. Mniej więcej od 1984 r. staje się ono dominujące, obecne jest szeroko w tekstach zarówno literackich, jak i publicystycznych, por.:

- (19) Niemal identyczne zdanie zanotowałem z wykładu psychiatry, którego wysłuchałem w domu kultury w Piotrkowie: „Ć p a się po to, żeby nie odbierać świata, jaki jest. Nasza młodzież, proszę państwa, jest smutna”.

(„Kino”, t. 18, 1984, books.google.pl)

- (20) Nienawidziłam matki [...] głucha i ślepa na ostrzeżenia i wyciągnięte ręce dobrych ludzi. Ja nie chciałam lepiej, chciałam ć p a ć, mocno żyć, być wolną.

(Maria Łopatkowa, *Nasz cudzy świat*, 1986, books.google.pl)

- (21) Gdybym nie zaczęła ć p a ć, to do tej pory albo już bym się dawno powiesiła, albo wylądowała w pasach. Dopiero po prochach zobojętniały mi te jego wieczne pretensje.

(Miroslaw Sokołowski, *Gady*, 1989, books.google.pl)

Warto nadmienić, że jeszcze w 1981 r. Klemens Stępniaak zanotował omawianą jednostkę jako element socjolektu hipisowsko-narkomańskiego (Stępniaak 1981: 320). Na podstawie tych wszystkich danych należy przyjąć, że przesunięcie semantyczne ‘jeść łąpczywie’ → ‘odurzać się narkotykami’ dokonało się w latach 70. w obrębie nowej (hipisi pojawiają się w Polsce pod koniec lat 60., por. Tracz 2014) odmiany środowiskowej polszczyzny. Była ona wówczas używana w bardzo wąskich kręgach, ale nie na tyle hermetycznych, by tłumacz amerykańskiego utworu literackiego nie mógł z niej zaczerpnąć potrzebnego ekwiwalentu tłumaczeniowego. W kolejnej dekadzie ów socjolektalny neosemantyzm został zapożyczony do języka ogólnego, a jego szybkie upowszechnienie się można wiązać z czynnikiem zewnątrzjęzykowym – epidemią narkomanii, która stała się w tamtym okresie palącym problemem społecznym. Ze względu na swoje pochodzenie omawiana jednostka leksykalna zachowała jednak bardzo silne negatywne zabarwienie emocjonalne.

5. Falsyfikacja kwalifikacji stylistycznej ekspresywizmów

Cyfrowe zbiory tekstów okazują się także przydatne przy ustalaniu rzeczywistych właściwości pragmatycznych jednostek leksykalnych, umożliwiając falsyfikację szeroko rozpowszechnionych sądów na temat ich nacechowania stylistycznego. Można to pokazać na przykładzie leksemu *obciach*. WSJP definiuje go jako ‘określenie wyrażające negatywną ocenę kogoś lub czegoś, co jest uznawane przez osobę mówiącą za brzydkie, niestosowne bądź żenujące’. Wskazany rzeczownik dość powszechnie zaliczany jest dziś do słownictwa młodzieżowego. Świadczy o tym fakt, że został odnotowany w takich opracowaniach leksykograficznych, jak *Nowy słownik gwary uczniowskiej* (Zgólkowa (red.) 2004) czy też *Słownik slangu młodzieżowego* (Czeszewski 2001). Tymczasem jednostka ta pojawia się już w SJPDor z kwalifikatorem *wiech*. (=wiechowe), co świadczy o zaliczeniu jej do gwary miejskiej. O *obciachu* wspomina również w artykule poświęconym współczesnemu (w latach 50.) słownictwu środowiskowo-emocjonalnemu Danuta Buttler (1959: 75). Badaczka uznała go za jeden z przykładów „nowotworów emocjonalnych”, których ekspresywność wynika z kontrastu między potocznym znaczeniem a specyficzną budową słowotwórczą – formacje postwerbalne uznaje autorka za typowy element terminologii specjalnej. Dawność słowa pozwala powątpiewać w to, by we współczesnej polszczyźnie miało ono specyficznie młodzieżowy charakter, leksyka używana przez młode pokolenie cechuje się bowiem wyjątkową zmiennością i nieustannym odświeżaniem zasobów.

Analiza danych cyfrowych potwierdziła nasze wątpliwości. Przynajmniej wykazała, że omawiane słowo pojawiło się w pisanej odmianie polszczyzny już w okresie przedwojennym, jego najstarsze poświadczenie pochodzi z roku 1929:

(22) KUCHARKA:

Gdy słońce na durślak fest rozpęta krzaki,
To zara me państwo wali na letniaki,
Na słońcu się smażyć, na piachu się parzyć,
W domu panią jestem ja...
Rypcium... rachu... ciachu...
Sztynno, bez o b c i a c h u ...
W domu panią jestem ja!...

(Stefan Domański, *Flirt majowy*, „Trubadur Warszawy”, nr 23, 1929, polona.pl)

Można przytoczyć również inne przykłady pochodzące ze schyłku lat 20. i początku 30.:

- (23) Każdy robił tam cwaniaka,
Aż się nagle robi draka,
Felek wrzeszczy zjeżdżaj łąchu,
Ignac wali bez obciachu,
Hop, siup, siup!

(Alfred Currin-Kurka, *Zabawa na Woli*, „Trubadur Warszawy”, nr 28, 1929, polona.pl)

- (24) Jak wrócił w nocy Karol, to baba mu zrobiła rewizję osobistą. Karol był w strachu, bo miał na sercu, w tylnej kieszeni od portek, fotografię tej podfruwajki. Nawalał się, by go żona przed „trybunał teściowy” nie postawiła, więc bez obciachu zwinął w kulkę fotografię i zjadł. Ta fotografia stanęła mu w poprzek w żondołku i w żaden sposób jej nie mógł strawić.

(Stefan Domański, *Wicek Pipszok*, „Trubadur Warszawy”, nr 31, 1932, polona.pl)

Należy zauważyć jednak, że przedwojenne użycia analizowanego słowa pochodzą wyłącznie z jednego, i to dość specyficznego, tytułu prasowego – „Trubadura Warszawy”, pisma o charakterze kabaretowym, publikującego teksty humorystyczne inspirowane realiami i folklorem stolicy. We wszystkich przykładach *obciach* stanowi element stylizacji językowej. Fakty te wskazują dość jednoznacznie, że omawiany wyraz wywodzi się z gwary miejskiej Warszawy, a w ówczesnym języku pisanym pojawiał się wyłącznie jako leksykalny ewokant mowy jej mieszkańców, reprezentowanych przez charakterystyczne typy ludzkie (kucharka, warszawski cwaniak), oraz właściwego dla nich stosunku do rzeczywistości. Podobną funkcję stylizacyjną pełnią najstarsze znalezione przeze mnie użycia powojenne, pochodzące z powieści dotyczących realiów warszawskich, takich jak *Zły* Leopolda Tyrmanda:

- (25) Jeszcze dziś będziesz na mieście, bez obciachu. Mucha nie siada.

(Leopold Tyrmand, *Zły*, 1956, books.google.pl)

czy też *Kolumbowie – rocznik 20* Romana Bratnego:

- (26) Olo opuścił Zygmunta i trzyma teraz za szyję Kolumba.
– Machabeusz, czego się złościsz, synku, ale obciach z tym twoim pseudem – śmieje się serdecznie, chowając twarz na piersiach przyjaciela, aż Kolumb stęka,

a dotychczas wyniośle obojętna na jego wezwania Niteczka spada na Ola jak kotka broniąca małego.

– ..ale ob c i a c h – bełkocze Olo, choć Kolumb-Machabeusz skręca się ze złości.

(NKJP: Roman Bratny, *Kolumbowie – rocznik 20*, 1957)

Jednak stosunkowo szybko pojawiają się po wojnie użycia, które rzeczywiście mogą sugerować, że omawiane słowo stało się popularne wśród młodzieży. Najstarszym i szczególnie wyrazistym przykładem jest satyryczna sztuka Andrzeja Wydrzyńskiego *Uczta morderców*:

(27) ROBERT: Helo, hawajka! (rozgląda się po pokoju) Osk! (gwizdze) Ale ob c i a c h.
Maksymalny.

KORNELIA: Co to jest maksymalny ob c i a c h?

ALICJA: Wielka kompromitacja.

(Andrzej Wydrzyński, *Uczta morderców*, 1960, books.google.pl)

Robert, który używa omawianej jednostki leksykalnej, to młody chuligan. Jego wypowiedziom, kontrastującym mocno z literacką polszczyzną innych bohaterów (wykształconych i zamożnych mieszczan), autor starał się nadać wyraziste cechy socjolektalne. Całościowa kreacja bohatera wskazuje, że mają one imitować język ówczesnej młodzieży, i to związanej raczej z marginesem społecznym.

Na młodzieżowy charakter leksemu może wskazywać również jego wystąpienie w zbiorze prac pamiętnikarskich młodych Polaków:

(28) Kolega Jakubik, gdy tylko zobaczył, że w zupie coś widać, za baniak i w kąt pod ścianę, bo tam czuł się bezpieczniej i nie robił takiego ob c i a c h u.

(*Młode pokolenie Polski Ludowej. Pamiętniki i studia*, 1964, books.google.pl)

a także w powieści przeznaczonej dla młodych czytelników i dotyczącej ich pokolenia:

(29) Jak bym jej teraz nie wyrzuciła, zadławiłaby mnie chyba. W jednym dniu straciłam przyjaciółkę i chłopaka. Żaneta i Bogdan ryzykowali narażenie się na ob c i a c h klasy.

(Maria Łopatkowska, *Nasz cudzy świat*, 1986, books.google.pl)

Równolegle z tego rodzaju użyciami pojawiają się jednak przykłady niemające nic wspólnego z kręgami młodzieżowymi i stylizacją na ich język. Są wśród nich takie, które przywołują wprost, za pomocą informacji meta-językowej, asocjacje przestępcze:

- (30) Zastanawia się, mówi powoli: – No tak, oni nie lubią tego wszystkiego, tego frajerstwa, może być kocka z tego powodu, i o b c i a c h dla ciebie. – (Ha, ha, on już używa tego cwaniacko-złodziejskiego żargonu [...]).

(„Twórczość”, t. 18, 1962, books.google.pl)

Inne ewokują jednak wyłącznie kolokwialność, por.:

- (31) Umówiliśmy się, że będziemy udawać walkę i bez o b c i a c h u ambicyjnego udawaliśmy. Raz się żyje i jedną ma się przydzieloną na życie głowę. Ale były jaja! Znów zostałem wygwizdany, ale tym razem gwizdy mnie nie dotknęły.

(Edmund Pietryk, *Stały fragment gry*, 1981, books.google.pl)

- (32) Kadet „Rysiek” zwoływał partyzantów gwizdkiem, bez o b c i a c h u. Niosło się na pół wsi.

(Władysław Machejek, *Z wojny tej, wojny złej...*, 1978, books.google.pl)

Ostatni przykład wydaje się szczególnie charakterystyczny, książka Władysława Machejka bowiem to zbiór wspomnień wojennych, co może sugerować, że słowo było autorowi znane przynajmniej z czasów II wojny światowej, a nie pochodził on z Warszawy, lecz z Małopolski. Wskazywałoby to na bardzo wczesne przeniknięcie omawianego wyrazu do innych regionów Polski.

Przegląd materiału pokazuje, że słowo *obciach*, nawet jeżeli w pewnych okresach cieszyło się szczególną popularnością wśród młodych ludzi, nigdy nie było jednostką specyficzną młodzieżową. Leksem ten, pierwotnie należący do gwary warszawskiej oraz spokrewnionego z nią żargonu przestępczego, bardzo szybko stał się kolokwializmem rozpowszechnionym w różnych środowiskach.

6. Zakończenie

W niniejszym artykule przedstawiłem tylko kilka możliwości wykorzystania danych cyfrowych w analizie diachronicznej leksyki ekspresywnej. Nie twier-

dzę przy tym, że jest to lista zamknięta, z pewnością można ją rozszerzyć. Dobór procedur analitycznych jest w znacznej mierze subiektywny, stanowi odzwierciedlenie doświadczeń badawczych autora, związanych m.in. z pracami nad chronologizacją słownictwa polskiego na potrzeby *Wielkiego słownika języka polskiego PAN* (WSJP). Wydaje się jednak, że omówione przykłady w przekonujący sposób pokazały, że internetowe kolekcje tekstów, szczególnie podawanych *in crudo* (biblioteki cyfrowe), a nie w postaci przetworzonej za pomocą narzędzi inżynierii lingwistycznej (korpusy), są już w tej chwili nieodzowną pomocą w eksploracji słownictwa rzadkiego i nacechowanego, takiego jak ekspresywizmy. Dotyczy to szczególnie jego nowszej warstwy, dlatego też do analizy wybrane zostały jednostki używane współcześnie i uchodzące za stosunkowo świeże. Dokładniejsze rozpoznanie zawartości omówionych zasobów przez lingwistów z pewnością ujawniłoby ich użyteczność w badaniach nad innymi klasami leksyki polskiej.

Literatura

- Burkacka I., 2001, *Porównawcza analiza gniazdowa wybranych leksemów neutralnych i nacechowanych ekspresywnie*, Warszawa: Elipsa.
- Buttler D., 1959, Słownictwo środowiskowo-emocjonalne we współczesnej polszczyźnie, „*Poradnik Językowy*”, z. 1–2, s. 68–78.
- Buttler D., 1979, Powojenne ekspresywizmy polskie, „*Prace Filologiczne*”, t. 29, s. 85–90.
- Derwojedowa M., 2017, Językoznawstwo komputerowe i inżynieria lingwistyczna po 1989 roku, „*Język Polski*”, R. XCVII, z. 1, s. 7–17.
- Grabias S., 1981, *O ekspresywności języka. Ekspresja a słowotwórstwo*, Lublin: Wydawnictwo Lubelskie.
- Jochym-Kuszlikowa L., 2002, Argotyzmy pochodzenia rosyjskiego we współczesnym języku polskim, „*Annales Academiae Paedagogicae Cracoviensis. Studia Linguistica*”, nr 1, s. 97–110.
- Kreja B., 1969, *Słowotwórstwo rzeczowników ekspresywnych w języku polskim. Formacje na -ik, -k-, -isko i -ina*, Gdańsk: Gdańskie Towarzystwo Naukowe.
- Król M., Derwojedowa M., Górski R. L., Gruszczyński W., Opaliński K., Potoniec P., Woliński M., Kieraś W., Eder M., 2019, Narodowy Korpus Diachroniczny Polszczyzny. Projekt, „*Język Polski*”, R. XLIX, z. 1, s. 92–101.
- Książek-Bryłowa W., 2001, O potoczności „Ogrodu fraszek” Wacława Potockiego. – *Poznańskie spotkania językoznawcze*, t. 7, red. Z. Krążyńska, Z. Zagórski, Poznań: Poznańskie Towarzystwo Przyjaciół Nauk, s. 87–97.

- Lee H., 2019, *15 Years of Google Books*, www.blog.google (dostęp: 16.04.2021).
- Milejkowska H., 1966, Kilka uwag o rusycyzmach w tzw. wiechu i w języku potocznym, „*Poradnik Językowy*”, nr 6, s. 256–260.
- Ogrodniczuk M., 2017, Lingwistyka komputerowa dla języka polskiego: dziś i jutro, „*Język Polski*”, R. XCVII, z. 1, s. 18–28.
- Pałka P., Kwaśnicka-Janowicz A., 2017, *Przewodnik po elektrycznych zasobach językowych (słowniki, kartoteki, korpusy, kompendia)*, Kraków: Towarzystwo Miłośników Języka Polskiego, tmjp.pl/images/pdf/przewodnik_po_elektrycznych_zasobach.pdf (dostęp: 16.04.2021).
- Przepiórkowski A., Bańko M., Górski R. L., Lewandowska-Tomaszczyk B. (red.), 2012, *Narodowy Korpus Języka Polskiego*, Warszawa: Wydawnictwo Naukowe PWN.
- Rejter A., 2006, *Leksyka ekspresywna w historii języka polskiego. Kulturowo-komunikacyjne konteksty potoczności*, Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Stępnia K., 1981, Słownictwo hipisowsko-narkomańskie w Polsce, „*Prace Filologiczne*”, t. 30, s. 317–336.
- Tracz B., 2014, *Hippiesi, kudłacze, chwasty. Hipisi w Polsce w latach 1967–1975*, Katowice–Kraków: Wydawnictwo Libron.
- Walczak B., 1992, Siedemnastowieczny styl niski (na przykładzie „Liber chamorum” Waleriana Nekandy Treпки). – *Odmiany polszczyzny XVII wieku*, red. H. Wiśniewska, Cz. Kosyl, Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej w Lublinie, s. 181–189.
- Wawrzyńczyk J., 2010, *Inny „Doroszewski”*, Łask: Leksem.
- Wierchoń P., 2008, *Fotodokumentacja. Chronologizacja. Emendacja. Teoria i praktyka weryfikacji materiału leksykalnego w badaniach lingwistycznych*, Poznań: Sorus.
- Żmigrodzki P., Bańko M., Batko-Tokarz B., Bobrowski J., Czelakowska A., Grochowski M., Przybylska R., Waniakowa J., Węgrzynek K. (red.), 2018, *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*, Kraków: Instytut Języka Polskiego PAN/Libron.

Biblioteki cyfrowe, korpusy, słowniki

- Czeszewski M., 2001, *Słownik slangu młodzieżowego*, Piła: Ekolog.
- ESXVII: *Elektryczny słownik języka polskiego XVII i XVIII wieku*, red. W. Gruszczyński, sxvii.pl (dostęp: 16.04.2021).
- Google Books, books.google.pl (dostęp: 16.04.2021).
- ISJP, 2000, *Inny słownik języka polskiego*, t. 1–2, red. M. Bańko, Warszawa: Wydawnictwo Naukowe PWN.
- Kania S., 1995, *Słownik argotyzmów*, Warszawa: Wiedza Powszechna.
- KJP XIX: Korpus Języka Polskiego Lat 1830–1918, f19.uw.edu.pl (dostęp: 16.04.2021).

- Korba: Elektroniczny Korpus Tekstów Polskich z XVII i XVIII w., korba.edu.pl (dostęp: 16.04.2021).
- KP XVI: Korpus Polszczyzny XVI wieku, spxvi.edu.pl/korpus (dostęp: 16.04.2021).
- NFJP: Narodowy Fotokorpus Języka Polskiego, nfjp.pl (dostęp: 16.04.2021).
- NKJP: Narodowy Korpus Języka Polskiego, nkjp.pl (dostęp: 16.04.2021).
- NSP, 2006, *Nowe słownictwo polskie. Materiały z prasy lat 1993–2000*, cz. 4, red. T. Smółkowa, Kraków: Lexis.
- POLONA, www.polona.pl (dostęp: 16.04.2021).
- SPXVI: *Słownik polszczyzny XVI wieku. Edycja internetowa*, spxvi.edu.pl (dostęp: 16.04.2021).
- SJPDor, 1958–1969, *Słownik języka polskiego*, t. 1–11, red. W. Doroszewski, Warszawa: Państwowe Wydawnictwo Naukowe.
- SStp, 1953–2002, *Słownik staropolski*, t. 1–9, Wrocław: Ossolineum, t. 10–11, Kraków: Instytut Języka Polskiego PAN.
- SW, 1900–1927, *Słownik języka polskiego*, t. 1–8, red. J. Karłowicz, A. Kryński, W. Niedźwiedzki, Warszawa: nakładem prenumeratorów i Kasy im. Mianowskiego.
- USJP, 2003, *Uniwersalny słownik języka polskiego*, t. 1–4, red. S. Dubisz, Warszawa: Wydawnictwo Naukowe PWN.
- Wikiźródła, pl.wikisource.org (dostęp: 16.04.2021).
- WSJP: *Wielki słownik języka polskiego PAN*, red. P. Żmigrodzki, wsjp.pl (dostęp: 16.04.2021).
- Zgółkowa H. (red.), 2004, *Nowy słownik gwary uczniowskiej*, Wrocław: Europa.

The utility of new digital resources for diachronic analysis of expressive lexical units

The article deals with the possibilities of using new digital linguistic resources for studies into history of expressive vocabulary. In the first part, the author discusses the most important of them, focusing on text corpora and digital libraries, such as *Polona* and *Google Books*, showing their advantages and disadvantages. In the second part of the article, theoretical considerations are exemplified by the diachronic analysis of the evolution of several expressive lexemes (*zajebisty* [EN: *dope*], *zajebicie* [EN: *for the win*], *szantrapa* [EN: *shrew, hellcat*], *ćpać* [EN: *pop*], *obciach* [EN: *shame*]). It turns out that digital collections, especially libraries, contain a large number of unknown old texts, which may play a key role in dating rare lexical units as well as their semantic and pragmatic changes.

Key words: *digital linguistic, diachronic analysis, lexical chronology, expressive vocabulary, colloquial vocabulary*

