

RENATA PRZYBYLSKA  
UNIwersytet Jagielloński  
KRAKÓW

## BADANIA NAD POLSZCZYZNĄ MÓWIONĄ A LEKSYKOGRAFIA

W niniejszym opracowaniu zamierzam rozważyć wpływ badań nad polszczyzną XX w. mówioną na współczesną praktykę leksykograficzną w kontekście powstającego właśnie nowego *Wielkiego słownika języka polskiego* (WSJP) (por. m.in. Żmigrodzki, Bańko, Dunaj, Przybylska 2007). Rozwijane w latach 70. i 80. w wielu ośrodkach w Polsce badania nad językiem mówionym niestety straciły ostatnio na swej dynamice, jednak to, co zostało na tym polu osiągnięte, warto w jak największym zakresie przełożyć na praktykę tworzenia współczesnych słowników ogólnych języka polskiego.

Słownik powinien dokumentować stan języka w danym okresie historycznym, zwłaszcza odzwierciedlać zasób słownictwa z tego okresu i zdawać sprawę ze sposobu używania zarejestrowanych jednostek leksykalnych. Źródła do słownika powinny być oparte nie tylko na tekstach pisanych, ale i na mówionych, a może nawet przede wszystkim na mówionych, jako że każdy język naturalny podstawowo manifestuje się i istnieje w wersji mówionej, wersja pisana jest wtórna i dla manifestowania się języka niekonieczna. Nie trzeba w tym miejscu przypominać, że wiele języków dla nas egzotycznych do dziś istnieje przecież tylko i wyłącznie w wersji mówionej. Tymczasem dotychczasowe słowniki ogólne – właściwie prawie wszystkie – trzymają się w przeważającej mierze źródeł pisanych<sup>1</sup>. Nie oznacza to oczywiście, że język mówiony jest w tych słownikach całkowicie nieobecny. Pojawia się głównie dzięki źródłom pisanim reprezentującym przede wszystkim teksty dramatu oraz partie dialogowe tekstów literackich. Taka wąska reprezentacja „mówioności”

---

<sup>1</sup> Nawet słowniki z założenia rejestrujące leksykę obecną przede wszystkim w języku mówionym opierają się głównie na tekstach pisanych. Dotyczy to choćby *Słownika polszczyzny potocznej* Janusza Anusiewicza i Jacka Skawińskiego (1996) i *Słownika polskich leksemów potocznych* pod redakcją Władysława Lubasia (t. I–V). W *Słowniku polszczyzny potocznej* tak autorzy piszą o doborze źródeł: „Zestaw haseł został opracowany na podstawie ponad 50 000 wypisów z wybranych pozycji książkowych, prasy, radia i telewizji i przede wszystkim – co jest szczególnie ważne – ze sporządzanych w różnych sytuacjach komunikacyjnych – zapisów żywego języka mówionego” (Anusiewicz, Skawiński 1996: 7).

jest oczywiście niezadowolająca. Co w tej kwestii można zmienić w przygotowywanym właśnie WSJP?

Po pierwsze, należy liczyć na to, że skoro WSJP będzie korzystał z Narodowego Korpusu Języka Polskiego (NKJP), to twórcy tegoż korpusu nie zapomną o języku mówionym. Z dotychczasowych informacji na temat struktury korpusu wynika niestety, że nadal dominują w nim teksty pisane<sup>2</sup>. Uwzględniane są właściwie jedynie te teksty mówione, które są dostępne już w wersji pisanej, a dokładniej w wersji elektronicznej<sup>3</sup>. W grę wchodzi na przykład stenogramy przemówień sejmowych. Choć lepsze to niż nic, to jednak ten typ źródeł trudno uznać za zadowalający z punktu widzenia celu, jakim jest wiarygodna słownikowa dokumentacja żywego języka mówionego. Po pierwsze stenogramy rejestrują tylko jedną z odmian stylistycznych języka mówionego – odmianę przynajmniej z założenia oficjalną, podczas gdy istotne osobliwości żywej mowy występują głównie w odmianie nieoficjalnej, spontanicznej, używanej na co dzień w niezobowiązujących sytuacjach potocznych (por. m.in. Przybylska 1994). Po drugie – stenogramy zazwyczaj nie stanowią wiernego zapisu żywej mowy, wiele elementów dla mówionej wersji charakterystycznych jest w nich po prostu celowo pomijanych<sup>4</sup>. W tej sytuacji należałoby koniecznie uzupełnić korpus o autentyczne źródła mówione, zwłaszcza pochodzące z odmiany nieoficjalnej. Można by wykorzystał istniejące już, wydane w druku, wybory tekstów języka mówionego, które powstawały w różnych ośrodkach badawczych w latach 80. XX w. i później. Mam na myśli Kraków, Łódź, Katowice, Poznań (por. Dunaj 1979; Lubaś 1978, 1980; Zgólkowa 1983). Można by też poddać dygitalizacji zarchiwizowane w tych ośrodkach istniejące nagrania języka mówionego, które nie weszły do opublikowanych wyborów tekstów. Nie jest to zadanie łatwe. Ponadto należałoby w sposób planowy, oparty na wypracowanych swego czasu założeniach metodologicznych (por. m.in. Kamińska 1979; Awdiejew, Labocha, Rudek 1980; Dunaj 1981) zacząć tworzyć korpus polskiego języka mówionego, który jako swoisty subkorpus wszedłby do NKJP, ale też sam w sobie stanowiłby cenne źródło do wszelkich badań nad językiem mówionym właśnie. Jest to w moim przekonaniu bardzo pilne zadanie badawcze, które musiałoby przyjąć formę odrębnego projektu, wymagającego oczywiście sporych nakładów finansowych i organizacyjnych<sup>5</sup>. Jeszcze innym zadaniem

<sup>2</sup> W referacie *Narodowy korpus języka polskiego – stan prac* przedstawionym na LXVI Zjeździe Polskiego Towarzystwa Językoznawczego w Opolu w 2008 r. autorzy: Rafał Górski, Barbara Lewandowska-Tomaszczyk, Adam Przepiórkowski, Marek Łaziński podali informację, że powstający obecnie Narodowy Korpus Języka Polskiego będzie obejmował 1 mld wyrazów, w tym 30 mln (za tym tylko 3% całości) będą stanowić użycia wyekscerpowane z tekstów mówionych, w tym także pochodzących z radia i telewizji.

<sup>3</sup> Takie postępowanie wynika zapewne głównie ze względów praktycznych i ekonomicznych. Łatwiej jest bowiem włączyć do korpusu teksty istniejące już w wersji elektronicznej niż najpierw zgromadzić teksty mówione, a potem je przetworzyć (zdygitalizować) na potrzeby korpusu.

<sup>4</sup> O niepełnej wiarygodności stenogramów przekonują między innymi badania prowadzone przez Magdalenę Smoczyńską na Wydziale Polonistyki UJ, dotyczące działalności komisji sejmowej w sprawie Lwa Rywina. Porównanie istniejących stenogramów z prac tej komisji z nagraniami telewizyjnymi ujawniło liczne rozbieżności między kształtem językowym autentycznych wypowiedzi a ich stenogramowym zapisem.

<sup>5</sup> O potrzebie włączenia w szerszym zakresie do słownika języka ogólnego słownictwa języka mówionego byli już przekonani projektodawcy słownika, który miał powstać w IJP PAN w latach 80.,

byłoby stworzenie korpusu tekstów mówionych z okresów wcześniejszych szeroko rozumianej współczesności, czyli od 1945 roku do połowy lat 70. XX w. Tu prawdopodobnie jedynym dostępnym źródłem byłyby nagrania dźwiękowe z archiwum polskiego radia i telewizji, jako że z tego okresu nie ma zbiorów celowo zaprojektowanych do badań naukowych żywego języka mówionego.

Dotychczasowe badania nad językiem mówionym przyniosły ważne wyniki teoretyczne, które należałoby uwzględnić w opisie słownikowym polszczyzny. Przede wszystkim chodzi o zarejestrowanie samego tego materiału leksykalnego, który jest charakterystyczny prawie wyłącznie dla polszczyzny mówionej. Już na samym etapie wyodrębniania jednostek leksykalnych można wskazać hasła z języka mówionego, których w dawnych słownikach nie było. Część z nich wymienili swego czasu w swym artykule Mirosław Bańko i Anna Kłosińska (1994), np. *znaczy/znaczy się* w funkcji wykładnika korekty w języku mówionym, inne zostały już dostrzeżone i ujęte na przykład w SWJP. Na podkreślenie zasługuje wyodrębnienie w tymże słowniku trzech klas części mowy, szczególnie często obecnych w języku mówionym: tak zwanych modulantów (np. *głównie, niestety*), operatorów metatekstowych (np. *no, że tak powiem, prawda*) i dopowiedzeń (np. *naturalnie, oczywiście, owszem*), które w dawniejszych słownikach albo były pozbawione jakiejkolwiek kwalifikacji gramatycznej, albo były włączone do niejasno zdefiniowanej klasy tzw. partykuł, albo też znalazły się co prawda w słowniku, lecz zanurzone głębiej w różnych miejscach artykułu hasłowego.

Zresztą zarówno co do wyodrębniania, jak i klasyfikacji wielu jednostek leksykalnych tego rodzaju trwają dyskusje, przy czym największy wkład w tym zakresie wnoszą obecnie prace Macieja Grochowskiego (m.in. 1986, 1997, 2008) i Jadwigi Wajszczuk (2005). Ponieważ opis tych jednostek zarówno gramatyczny, jak i semantyczny, jest właściwie *in statu nascendi*, można się spodziewać znacznej rozbudowy liczebnej tych klas gramatycznych, jak i istotnych nowych rozstrzygnięć teoretycznych. Trudno jednak zaprzeczyć, że wielu jednostek tego rodzaju badacze nie zauważyliby, gdyby nie poważnie potraktowane jako źródło do opisu leksyko-graficznego języka wiarygodne teksty mówione<sup>6</sup>. Tym tekstom między innymi pod takim kątem jako jeden z pierwszych przyjrzał się swego czasu Kazimierz Ożóg (1990) w swej pracy o metaleksykonie polszczyzny mówionej.

---

o czym pisał między innymi Alfred Zaręba (1988). Postulował on na przykład przeprowadzenie w tym celu wyrwykowych, choć ściśle sprecyzowanych badań własnych, godząc się jednak na rozwiązanie polowiczne: „słownictwo języka mówionego w pewnym stopniu można zarejestrować przez teksty pisane, które się na mówionym opierają. Nie jest to rozwiązanie najlepsze, ponieważ w takim [...] wypadku mielibyśmy do czynienia nie z żywym tekstem mówionym, ale wtórnie zarejestrowanym” (Zaręba 1988: 43).

<sup>6</sup> W pracach M. Grochowskiego (m.in. 1986, 1997) także w przykładowych artykułach hasłowych dla tych jednostek egzemplifikacje użyć w poszczególnych znaczeniach często wyraźnie odzwierciedlają uzus w języku mówionym, zwłaszcza tam, gdzie jednostka ujawnia swe funkcje gramatyczne i semantyczne właściwie tylko w żywym dialogu, na przykład hasło *akurat 3* zakwalifikowane jako dopowiedzenie w użyciach typu: *Czy Adam wrócił? – Akurat!, Dostałeś pewnie nagrodę? – Akurat!, Chociaż wstąpił po ciebie? – Akurat!* (por. Grochowski 1997: 50–51)

Studia nad językiem mówionym przyczyniły się też do rozwoju badań pragmatycznojęzykowych. Dzięki nim do repertuaru jednostek języka ujętych jako hasła w słowniku dołączyły takie, które dziś nazwalibyśmy frazematami, np. formuły etykietalne powitań, pożegnań, podziękowań, przeprosin itd. Zwrócono też uwagę na potrzebę rozbudowania opisu słownikowego wielu jednostek o informacje właśnie pragmatycznojęzykowe, uwzględniające to kto, do kogo, w jakiej sytuacji i po co tak mówi. Przykładowo dla hasła *proszę* w ISJP, który częściowo uwzględnił już zjawiska związane z językiem mówionym, i to stanowi jego wielką zaletę, dla podhasła *proszę* opisano następujące konteksty pragmatycznojęzykowe: „mówimy «proszę», kiedy” a) „podajemy lub pokazujemy coś komuś” [...], b) „zachęcamy kogoś, aby coś zrobił, np. wszedł dokądś lub usiadł gdzieś [...]”, c) „wyrażamy zgodę na czyjąś propozycję lub gotowość zrobienia czegoś [...]”, d) „ktoś podziękował nam za przysługę [...]”, e) „ktoś zapukał do naszych drzwi i chcemy mu dać znać, aby wszedł [...]”, f) „nie dosłyszeliśmy tego, co powiedział rozmówca, i chcielibyśmy, aby to powtórzył [...]”. Wielu z tych informacji pragmatycznych we wcześniejszych słownikach nie było. Dziś na przykład musielibyśmy dołączyć do formuł etykietalnych modne wyrażenie *Milego dnia! czy W czym mogę pomóc?*, oba zapewne upowszechnione jako kalki z języka angielskiego.

Przyjęcie tekstów mówionych jako równoprawnego źródła do słownika w praktyce rodzi liczne szczegółowe problemy wymagające konsekwentnego rozstrzygnięcia. Niektóre z nich zostaną tu pokrótce zasygnalizowane.

1. Czy w WSJP ująć jako osobne hasła potoczne, kolokwialne warianty fonetyczne i w konsekwencji pisowniane niektórych wyrazów, np. *se* zamiast *sobie*, *psorka* zamiast *profesorka*, *trza* zamiast *trzeba*? Czy może te formy umieścić tylko jako warianty wymowy w wersji potocznej (wtedy jednak pominęlibyśmy informację, że w takiej postaci te jednostki są nie tylko wymawiane, ale i zapisywane)? Jeśli tak, to jak daleko należy się posunąć, gdzie postawić granicę między formami obocznymi, potocznymi, które zasługują na rejestrację, a tymi, które nie będą odnotowane? Czy zdawać też sprawę z wymowy typu: *uniwerstet*, *orginalny*, *wogle*, czy już nie? Przeglądając dotychczasowe słowniki odnosi się wrażenie, że rejestrowano głównie te warianty (w zakresie wymowy, jak i pisowni tę wymowę odzwierciedlającej), które znalazły się w tekstach pisanych, przede wszystkim w partiach stylizowanych właśnie na żywą potoczną mowę.

2. W ogóle podejście do języka w leksykografii było i nadal jest bardzo „graficzne”, to znaczy status obiektów leksykograficznych najłatwiej zyskują te, które mają utrwaloną postać pisaną, trudniej wchodzi do słownika jednostki właściwie żyjące tylko w mowie, jak np. seria *eska*, *emka*, *elka*, *ikselka* / *S-ka*, *M-ka*, *L-ka*, *XL-ka* (?) w znaczeniu ‘sztuka odzieży w rozmiarze S (ang. *small*) lub rozmiar S, w rozmiarze L (ang. *large*), w rozmiarze XL (ang. *extra large*)’ itd., które to wyrazy właściwie nie mają chyba ustalonej pisowni, bo pojawiają się prawie tylko i wyłącznie w mowie. Ten problem ma zresztą swą drugą stronę, rodzi się bowiem pytanie, czy status pełnoprawnych jednostek języka mają te, które istnieją na przykład tylko jako napisy, właściwie nigdy nie realizują się w mowie. Dotyczy to wielu oznaczeń na towarach, przykładowo kolory damskich rajstop są oznaczane słowami: *nero*, *fumo* itd. (z języka włoskiego), ale czy mamy prawo twierdzić, że są to zapożyczenia zaadaptowane

w polszczyźnie, skoro w kodzie mówionym nikt z Polaków takich wyrazów nie używa? Na marginesie zwrócę uwagę na to, że na przykład Elżbieta Mańczak-Wohlfeld (2006) na sporządzonej przez siebie liście zapożyczeń angielskich umieściła między innymi jednostkę *VSOP* [napis umieszczany na butelkach koniaku wskazujący na jego najwyższą jakość, powstały od ang. wyrażenia *V(ery) (S)pecial (O)ld (P)ale brandy* (Mańczak-Wohlfeld 2006: 166)], której postać mówiona chyba po prostu nie istnieje. Podobnie jednostka *L* (od ang. *learner driver*) umieszczana na samochodach szkół nauki jazdy, która istnieje chyba w naszej świadomości jako byt tylko graficzny, pewien semiotyczny symbol, a nie jako wyraz.

3. Problem przekładu języka z mówionego na pisany i *vice versa* z pisanego na mówiony łączy się też z niejasnym statusem leksykograficznym i w ogóle językowym tych elementów kodu mówionego, które zalicza się do tzw. parajęzyka. Chodzi o sygnały fonetyczne o różnych funkcjach pragmatycznojęzykowych: *hm* (odnotowane w ISJP), *mhm* (nie ujęte w ISJP), *a*, *aaa*, *e* (w ISJP uznane za wykrzykniki), tzw. znaki retardacji lub „jęki namysłu”: *y*, *yy...*, *ee...* Te elementy zazwyczaj giną jako pierwsze w procesie transkrypcji tekstów mówionych na pismo. Ocalały są jedynie takie, dla których w tekstach pisanych, szczególnie literackich, ustaliła się już jakaś graficzna konwencja, np. *mhm*.

4. Pytając o granice mówioności w słowniku, pytamy o granice potoczności, ale i o granice między formami poprawnymi a niepoprawnymi.

Problem: odnotowywać czy nie to, co jest w języku, ale uchodzi za niepoprawne, dotyczy nie tylko samych haseł i ich postaci, ale form fleksyjnych pewnych wyrazów hasłowych. Czy i w jakim zakresie rejestrować, co prawda, substandardowe, uznawane za niepoprawne, ale jednak bardzo rozpowszechnione konstrukcje fleksyjne? Dotyczy to na przykład form imiesłowowych od czasownika *potrafić* typu: *potrafiąc* (!), *potrafiący* (!). Tych form na przykład ISJP w ogóle nie odnotowuje. Ale już forma *se* została w ISJP odnotowana z odesłaniem do *się* i tam opatrzona komentarzem typu: *w celowniku, zwłaszcza w znaczeniu 1. potocznie bywa używana forma se*. Jeśli słownik ma mieć charakter przede wszystkim dokumentacyjny, a dopiero w drugim kroku normatywny, to wydaje się, że nie powinien takich użyć pomijając, choć będą one oczywiście opatrzone komentarzem normatywnym typu „uważane za niepoprawne”<sup>7</sup>. W kwestii wariantywności fleksyjnej w grę wchodzi też zróżnicowanie regionalne niektórych form, spośród których jedne są piętnowane jako niepoprawne, inne dopuszczalne na równych prawach, np. *swędzieć* i *swędzić* – obie są poprawne, choć jedna kwalifikowana jako regionalna. Wariantywność fleksyjna języka zdeterminowana regionalnie właściwie ujawnia się wyłącznie w języku mówionym mieszkańców danego regionu. Wydaje się mało prawdopodobne, żeby w WSJP udało się autorom ten typ wariantywności zarejestrować, gdyż wymagałoby to szeroko zakrojonych badań terenowych. Takie badania jednak warto w przyszłości podjąć.

Powyższe punkty oczywiście nie wyczerpują kwestii związanych z rejestrowaniem w słowniku ogólnym języka polskiego zjawisk właściwych polszczyźnie przede

<sup>7</sup> Informacja normatywna w WSJP będzie zgodna z orzeczeniami normatywnymi w *Wielkim słowniku poprawnej polszczyzny PWN* pod redakcją Andrzeja Markowskiego (2004).

wszystkim w odmianie mówionej. Świadczą jednak o potrzebie zintensyfikowania badań nad żywym językiem mówionym, zwłaszcza tym używanym w sytuacjach nieoficjalnych w bezpośrednich kontaktach językowych. Żywa mowa niewątpliwie najtrudniej poddaje się metodycznej rejestracji, jak i naukowemu opisowi, nie znaczy to jednak, że ten obszar badań ma być zaniedbany. Leksykografia współczesna nie może rezygnować z opisu tego, co dla tej, w końcu prymarnej odmiany polszczyzny, jest charakterystyczne. Postęp w zakresie technicznych możliwości rejestrowania języka mówionego powinien współczesnym badaczom wymiennie ułatwić pracę. Można więc oczekiwać, że polszczyzna mówiona przestanie być „ubogą krewną” polszczyzny pisanej zarówno w naukowych opisach ogólnych języka, jak i w nowoczesnych opisach leksykograficznych.

## Literatura

- ANUSIEWICZ J., SKAWIŃSKI J., 1996, *Słownik polszczyzny potocznej*, Warszawa–Wrocław.
- AWDIEJEW A., LABOCHA J., RUDEK K., 1980, *O typologii tekstów języka mówionego*, „Polonica” VI, s. 181–187.
- BAŃKO M., KŁOSIŃSKA A., 1994, *Polszczyzna mówiona, nieobecna w słownikach*, [w:] Z. Kurzowa, W. Śliwiński (red.), *Współczesna polszczyzna mówiona w odmianie opracowanej (oficjalnej)*, Kraków, s. 89–96.
- DUNAJ B. (red.), 1979, *Wybór tekstów języka mówionego mieszkańców Krakowa*, Kraków.
- DUNAJ B. (red.), 1981, *Studia nad polszczyzną mówioną Krakowa. 1*, Kraków.
- GROCHOWSKI M., 1986, *Polskie partykuły. Składnia, semantyka, leksykografia*, Wrocław.
- GROCHOWSKI M., 1997, *Wyrażenia funkcyjne. Studium leksykograficzne*, Kraków.
- GROCHOWSKI M., 2008, *Założenia ogólne opisu tzw. wyrażeń funkcyjnych w Wielkim słowniku języka polskiego*, [w:] P. Żmigrodzki, R. Przybylska (red.), *Nowe studia leksykograficzne 2*, Kraków, s. 11–22.
- ISJP: M. Bańko, *Inny słownik języka polskiego*, Warszawa 2000.
- KAMIŃSKA M., 1979, *Założenia metodyczne badania polszczyzny mówionej w Łodzi i ich praktyczna realizacja*, „Socjolingwistyka” 2, s. 87–94.
- LUBAŚ W. (red.), 1978, *Teksty języka mówionego mieszkańców miast Górnego Śląska i Zagłębia. 1*, Katowice.
- LUBAŚ W. (red.), 1980, *Teksty języka mówionego mieszkańców miast Górnego Śląska i Zagłębia. 2*, cz. 1–2, Katowice.
- LUBAŚ W. (red.), 2001–2009, *Słownik polskich leksemów potocznych*, Kraków.
- MAŃCZAK-WOHLFELD E., 2006, *Angielsko-polskie kontakty językowe*, Kraków.
- MARKOWSKI A. (red.), 2004, *Wielki słownik poprawnej polszczyzny PWN*, Warszawa
- OZÓG K., 1990, *Leksykon metatekstowy współczesnej polszczyzny mówionej. Wybrane zagadnienia*, Kraków.
- PRZYBYLSKA R., 1994, *Porównanie struktury słownictwa częstego wypowiedzi mówionych w sytuacji oficjalnej i nieoficjalnej oraz tekstów pisanych w języku polskim*, [w:] Z. Kurzowa, W. Śliwiński (red.), *Współczesna polszczyzna mówiona w odmianie opracowanej (oficjalnej)*, Kraków, s. 97–106.
- SWJP: B. Dunaj (red.), *Słownik współczesnego języka polskiego*, Warszawa 1996.
- WAJSZCZUK J., 2005, *O metatekście*, Warszawa.

- ZARĘBA A., 1988, *Źródła do wielkiego słownika współczesnej polszczyzny. Uwagi ogólne*, [w:] W. Lubaś (red.), *Wokół słownika współczesnego języka polskiego*, Wrocław, s. 37–44.
- ZGÓLKOWA H., 1983, *Słownictwo współczesnej polszczyzny mówionej. Lista frekwencyjna i rangowa*, Poznań.
- ŻMIGRODZKI P., BAŃKO M., DUNAJ B., PRZYBYLSKA R., 2007, *Koncepcja Wielkiego słownika języka polskiego – przybliżenie drugie*, [w:] P. Żmigrodzki, R. Przybylska, (red.), *Nowe studia leksykograficzne*, Kraków, s. 9–21.

### **Studies in spoken Polish vs lexicography Summary**

The article discusses some selected problems involved in entering in the general Polish vocabulary of forms proper to spoken language. The author points to the need to create a corpus of contemporary spoken Polish which could make a credible source base for the comprehensive Polish dictionary now being developed.