

JAK POWSTAWAŁ *FOTOSUPLEMENT* DO SŁOWNIKA WARSZAWSKIEGO I CO Z TEGO WYSZŁO

PIOTR WIERZCHOŃ

1. Opis problemu

W czasie II wojny światowej naród polski pod okupacją niemiecką poniósł niewyobrażalne straty (m.in. utrata około 40% majątku narodowego sprzed roku 1939). Szczególnie bolesną stratą w ramach tych działań była grabież zabytków kultury polskiej (obrazy, rzeźby, fotografie, filmy itp.; por. Matelski, 2007). Zniszczenia objęły także zbiory biblioteczne i archiwalne.

Z językoznawczego punktu widzenia wielkim ubytkiem była utrata kartoteki *Suplementu* do Słownika warszawskiego, czyli *Słownika języka polskiego* pod red. A. Kryńskiego, J. Karłowicza i W. Niedźwiedzkiego. Lata 30. XX wieku były okresem dynamicznego rozwoju technologicznego i naukowego oraz przemian społecznych i politycznych, co skutkowało tworzeniem się nazw dla tych desygnatów i zjawisk (por. wielokrotne sondy w tym zakresie: Wawrzyńczyk, 1999; 2009; 2011).

Działanie podjęte w ramach ekscerpcji fotodokumentacyjnej miało doprowadzić do rekonstrukcji siatki haseł tego *Suplementu*. Założeniem projektu była zatem ekscerpcja 60 000 jednostek, które mogłyby uzupełnić Słownik warszawski. W wyniku przeprowadzonych prac uzyskano **ponad 77 tysięcy jednostek** z okresu 1930–1939. Dodatkowo wyekscerpowano **blisko 6000 jednostek** wielowrazowych z tego samego okresu. Daje to sumę około 83 tysięcy obiektów. Całość przedstawia się następująco:



Tomy rozpoczynają się od numeru 11 (do numeru 40), ponieważ *Fotosuplement* ukazuje się w serii B o zbiorczej nazwie: *Depozytorium leksykalne języka polskiego*. Tomy 1–10 (w czarnej obwolucie) ukazywały się od roku 2010, również w warszawskim wydawnictwie BEL Studio.

2. Preferencje działania

Ekscerpcja materiałów z lat 30. XX w. regulowana jest metodą działania wynikającą z:

- a) oceny realności wykonania danego zadania (nakład zasobów niezbędnych do wykonania zadania w określonym czasie; por. Kotarbiński, 1938; Węglarz, 1981),
- b) analizy stanu polskich zasobów leksykalnych, derywacyjnych, chronologicznych itp. (Wawrzyńczyk, 2000).

Poniżej przedyskutuje się te uwagi.

ad a).

Najbardziej realne obecnie wydaje się sporządzenie spisu derywatów afiksalnych, np. prefiksalnych (*bezalkoholowy, bezambicyjny, bezamoniakalny, bezamortyzacyjny, bezanonimowo, bezanonimowy, bezantenowy, bezantygenowy, bezapelacyjność* itd.), sufiksalnych, np. (*dymuszowaty* ← Adolf Dymśza, *albaniizacja, analizacja, angelizacja, anglikanizacja, archaizacja, archiwizacja, ateizacja* itd.), prefiksoidalnych (np. *superstabilizacja, superstacja, superstatek, superstruktura, superszlagier, superszlagierowy, superszwedzki, superszybkość, superteatr, supertechnik, supertempo, supertokarnia, supertomasyna, supertotalizm, supertrwały, supertytuł, superurbanizm, superwitamina, superwizować, superwychwalanie, superwyczyn, superwynałazek, superzeppelin, superżyd* itd.) złożzeń (*samochodokilometr*) oraz duoninów (np. *fachowiec-Niemiec, fachowiec-Polak, fachowiec-analfabeta, fachowiec-architekt, fachowiec-artysta, facho-*

wiec-astronom, fachowiec-automobilista, fachowiec-badacz, fachowiec-bankier, fachowiec-bohater, fachowiec-botanik, fachowiec-brakarz, fachowiec-buchalter, fachowiec-celnik, fachowiec-ceramik, fachowiec-cukiernik, fachowiec-czytelnik, fachowiec-drukarz, fachowiec-ekonomista, fachowiec-elektryk, fachowiec-erudyta, fachowiec-etnolog, fachowiec-filmowiec itd.). Mniej realne i mniej konieczne dla oceny ewolucji systemu leksykalno-derywacyjnego wydaje się obecnie ekscerpowanie neosemantyzmów, tj. takich jednostek, dla których w systemie językowym pojawiła się forma homograficzna o innym znaczeniu, np. *lustracja* ‘procedura ujawniania współpracowników służb bezpieczeństwa’, *momenty daw.* ‘przeważnie niejednoosobowe sceny o charakterze erotycznym, zazwyczaj formułowane w stonowanej – na tle pełni możliwości w tym zakresie – nieoryndarnej formie, zwłaszcza w dziele filmowym lub literackim niezorientowanym na ten typ afirmacji kontaktów cielesnych, aczkolwiek niewykluczającym przepłotu tego typu informacji z Amorem w tle’ itp.

ad b).

Najbardziej konieczne obecnie wydaje się sporządzenie rejestru derywatów dla okresu 1. poł. XX w. ze względu na obfitość materiału, najlepiej dla lat 30. ubiegłego wieku, ponieważ spis ten pozwoli uściślić szereg hipotez empirycznych wysuniętych w największych modelach leksykalno-derywacyjnych polszczyzny 2. poł. XX w.

Testowe ekscerpcje sporządzane dla materiału leksykalnego z lat 30. XX w. dla litery *a-* uświadamiały w roku 2010 (Wierzchoń, 2010c), że **trud rekonstrukcji polskiego przedwojennego zasobu leksykalnego** liczącego około 60 000 haseł nie jest pustosłownym dezyderatem. Zakończzone prace **wykazały, że jest to możliwe.**

3. Przebieg działania

Jako źródło danych wybrano obiekty zarchiwizowane w polskich bibliotekach cyfrowych wykorzystujących mechanizm *dLibra*. Mechanizm ten pozwala w ujednoczony sposób obserwować zasób bibliotek cyfrowych, dokonywać selekcji materiałów tam ujętych itp. Ostatecznie do zbioru tekstów, z których pobrano cytaty, weszło dla zbioru alfabetycznego *a–ż*, blisko 30 000 tekstów, z których pobrano średnio 2,5 cytatu. Dla 16 000 dokumentów pobrano tylko jeden cytat. Łącznie zatem, jak wspomniano, uzyskano około 77 tysięcy jednostek. Kryterium wyboru tekstów stanowiły następujące parametry:

- a) publikacja z okresu 1930–1939,
- b) względnie dobra jakość techniczna obrazu (dokumentu),
- c) preferencja dla czasopism.

Jeżeli chodzi o udział tekstów z poszczególnych bibliotek cyfrowych w końcowym wyniku, to przedstawia się on następująco dla najobszerniejszych pięciu wyników: ok. 14 000 dokumentów z Wielkopolskiej Biblioteki Cyfrowej, tyle samo z Małopolskiej Biblioteki Cyfrowej, ok. 9000 ze Śląskiej Biblioteki Cyfrowej, ok. 7000 z Biblioteki Cyfrowej UW oraz ok. 5000 z: Regionalia Ziemi Łódzkiej.

Dla każdego ekscerptu sporządzono dokumentację w postaci wycinka tekstowego, tak więc wszystkie jednostki uzyskały oryginalne potwierdzenie tekstowe.

Tak katolicyzacja Z. N. P., jak przemianowanie Stowarzyszenia z chrześcijańskiego na katolickie nie rozwiążą sprawy. Będą tylko wa-

kultura Cena 80 groszy

Poznań, niedziela 28 marca 1937 r.

Całość ujęto w 30 tomów o objętości ok. 750 stron na każdy tom. Udział poszczególnych haseł rozpoczynających się na daną literę przedstawia poniższa tabela:

a	2480	o	4127
b	1985	ó	7
c	1557	p	14000
ć	154	q	143
d	2695	r	3556
e	1557	s	8029
f	1379	ś	888
g	1882	t	3712
h	1241	u	1342
i	945	v	111
j	1358	w	6296
k	5062	y	27
l	1418	z	3562
ł	176	ż	14
m	2573	ź	1197
n	3579		

Aby uzyskać powyższy wynik, przeprowadzono następujące działania:

- a) dokonano preselekcji źródeł tekstowych w celu sporządzenia bazy materiałowej dla ekscerpcji,
- b) stworzono listę słów ujętych w tych dokumentach i przeprowadzono wstępną procedurę oczyszczania tekstu z rekordów, które wstępnie oceniono jako nie-słowa (np. ciągi łączników, ciągi z cyframi, ciągi 8 spółgłosek) itp.,
- c) dokonano automatycznej analizy listy słów ze względu na opracowane filtry ekscerpcyjne,
- d) wykonano bazę danych dla wyekscerpowanych jednostek,
- e) dokonano walidacji (ręcznej korekty) wyników uzyskanych w kroku c),
- f) przeprowadzono końcową edycję wyniku w postaci fotodokumentacji.

ad a). Dla przeprowadzenia badania konieczne było określenie zasobu tekstowego, który pozwoliłby na realizację systematycznego programu badań słów z okresu 1930–1939. Punktem docelowym było uzyskanie ok. 60 tysięcy jednostek. Optymalnym zbiorem danych polskojęzycznych dla tego typu badań okazały się teksty zgromadzone w systemie bibliotek cyfrowych – *dLibra*; do ekscerpcji weszły głównie czasopisma, np. *Codzienna Gazeta Handlowa*, *Czasopismo Geograficzne*, *Czasopismo Skarbowe*, *Droga*, *Fotograf Polski*, *Gazeta Handlowa*, *Głos Poranny*, *Goniec Wielkopolski*, *Kultura*, *Missye Katolickie*, *Moja Przyjaciółka*, *Myśl Narodowa*, *Nowiny Codzienne*, *Nowy Głos*, *Orędownik*, *Oświata Pozaszkolna*, *Polska Gospodarcza*, *Prosto z Mostu*, *Przegląd Artyleryjski*, *Przegląd Geograficzny*, *Przegląd Mechaniczny*, *Przegląd Pedagogiczny*, *Przegląd Piechoty*, *Samochód*, *Wychowanie Fizyczne* itp.

ad b). W ramach pracy wykorzystywano teksty źródłowe z lat 30. XX wieku. Zdecydowana większość tekstów ma postać cyfrową otrzymaną za pomocą skanowania i analizy oprogramowaniem OCR (ang. *optical character recognition*), które jest niedoskonałe. Na skutek tego niektóre teksty (zwłaszcza dzienniki) są zaszumione – pomiędzy słowami czy znakami rozpoznanymi prawidłowo znajdują się ciągi znaków będące efektem błędów programu rozpoznającego znaki, por.:

krelslerowle 1	krelsner 2	krelsowlęj 1
krelsleuerwehr 1	krelsnera 1	krelsowyc1 1
krelslg 1	krelsnia 1	krelsphusikus 1
krelslilem 1	krelso 1	krelsrun 1
krelslono 1	krelsoberinspektor 1	krelsrza 2
krelslowsklcj 1	krelsow 1	krelss 4
krelsm 1	krelsowa 1	krelssage 1
krelsmansa 1	krelsowe- 1	krelssi 1
krelsm1 1	krelsowej 1	krelssl 2
krelsnegostawu 2	krelsowicza 1	krelsspar- 1

ad c). Na tym etapie wszystkie teksty źródłowe zostały automatycznie prze-filtrowane za pomocą opracowanych filtrów, które pozwoliły na usunięcie błędnych fragmentów. Filtrami były głównie słownik PoliMorf oraz słownik, który powstał na podstawie tego zasobu poprzez sztuczne skonkatenowanie komponentu prefiksального z danym słowem z listy, przykładowo:

anty+

```
aaronowa aaronowy adj:sg:nom.voc:f:pos
aaronową aaronowy adj:sg:acc:f:pos
aaronową aaronowy adj:sg:inst:f:pos
aaronowe aaronowy adj:pl:acc:m2.m3.f.n1.n2.p2.p3:pos
aaronowe aaronowy adj:pl:nom.voc:m2.m3.f.n1.n2.p2.p3:pos
aaronowe aaronowy adj:sg:acc:n1.n2:pos
aaronowe aaronowy adj:sg:nom.voc:n1.n2:pos
aaronowego aaronowy adj:sg:acc:m1.m2:pos
aaronowego aaronowy adj:sg:gen:m1.m2.m3.n1.n2:pos
aaronowej aaronowy adj:sg:dat:f:pos
aaronowej aaronowy adj:sg:gen:f:pos
aaronowej aaronowy adj:sg:loc:f:pos
aaronowemu aaronowy adj:sg:dat:m1.m2.m3.n1.n2:pos
aaronowi aaronowy adj:pl:nom.voc:m1.pl:pos
aaronowo aaronowy adja
aaronowy aaronowy adj:sg:acc:m3:pos
```

Aby uzyskać listę potencjalnych derywatów, skonkatenowano listę prefiksów z listą wyrazów. Sporadycznie do zasobu włączano jednostki spoza listy filtracyjnej, wyodrębnione na podstawie manualnych analiz list słów z dokumentów.

ad d). Wykonano bazę danych, na którą składały się następujące informacje: słowo w postaci zlematyzowanej (hasłowej), słowo w postaci tekstowej (fleksyjnej), 3 obrazy, które zostały przyporządkowane do tego hasła, tzn. cytat, tytuł dokumentu, data dokumentu z dokładnością co najmniej roczną. Prócz tych danych utrzymywana jest informacja o rozmiarze (wielkości) 3 obrazów. Dążono do tego, aby każdą jednostkę ilustrować możliwie wczesnym potwierdzeniem z lat 30., przy czym istotnym kryterium wyboru była także dobra ilustracyjność kontekstu, jakość skanu itp.

ad e). Wszystkie hasła zostały poddane dwukrotnej korekcie, której istota polegała na wyeliminowaniu haseł, które mogłyby zawierać błąd literowy lub znajdować się w identycznej formie ortograficznej na liście haseł Słownika warszawskiego. Błędy literowe (z punktu widzenia dzisiejszej normy ortograficznej) o szczególnej wartości filologicznej, głównie te kodujące wariantywność rodzącej się w latach 30. polskiej normy ortograficznej, pozostawiono dla dalszych, osobnych analiz, por. (zapis po stronie lewej ujęto w wynik ekscerpcji fotodokumentacyjnej z lat 30., zapis po stronie prawej odzwierciedla zapis Słownika warszawskiego): *archidiecezja* – *archidjecezja*, *arcy-doskonały* – *arcydoskona-*

ty, arcy-pasterz – arcy-pasterz, asygnariusz – asygnarjusz, desmurgia – desmurgja, diakoński – djakoński, diametralnie – djametralnie, dwu-roczny – dwuroczny, ekstraktor – ekstraktor, głęboko-myślny – głębokomyślny, heliocentryczny – heljocentryczny, heliotropowy – heljotropowy, indygena – indygiena, interrex – interreks, kontubernia – kontubernja, kurierkowy – kurjerkowy, kwestionariusz – kwestjonarjusz, materialistyczny – materjalistyczny, megaohm – megaom, misterium – misterjum, monomaniak – monomanjak, notariat – notarjat, regencyjny – regiencyjny, registrator – regiestrator, rynoskopia – rynoskopja, różnokolorowy – różnokolorowy, różno-licowy – różnolicowy, różno-liczny – różnoliczny itd.

Pozostawiono także jednostki, które z językoznawczego punktu widzenia mogły wydawać się szczególnie interesujące, np. syntetyczne formy najwyższe przymiotnika, por.: *fantastyczniejszy, higieniczniejszy, hojniejszy, huczniejszy, inteligentniejszy, jednostajniejszy, jędrniejszy, najfantastyczniejszy, najordynarniejszy, powierzchniejszy, prawdopodobniejszy, produktywniejszy, przenajdziwniejszy, przenaigroźniejszy, przenaiprzedniejszy, przenaipłodniejszy, przenaiprozkoszniejszy, przenaipróznorodniejszy, przenaipsubtelniejszy, wielkoduszniejszy, wonniejszy, zaszczytniejszy, żółciejszy* itp.

ad f). Ostateczną fazą było przygotowanie publikacji, która ujmowałaby wszystkie wyekscerpowane hasła wraz z kontekstem fotodokumentacyjnym. Rozmiar materiału wymusił opracowanie 30 tomów po ok. 750 stron, co daje ok. 22 000 tekstu z fotodokumentacyjnymi poświadczeniami.

Reasumując, wniesiono do wiedzy o zasobach polszczyzny pierwszej połowy XX wieku ok. 80 000 jednostek z lat 30. ubiegłego wieku, które mogą stanowić uzupełnienie siatki haseł Słownika warszawskiego. Dodatkowo, w ramach wykonanych działań, uzyskano pewne interesujące informacje dotyczące struktury Słownika warszawskiego, jak i słownictwa z lat 30. XX wieku. Wyekscerpowane słowa ułożyły się w pewne grupy:

- a) tematyczne (medycyna: *acetonemia, adenopatia, angioneurotyczny, antyserum, apendektomia, auto-infekcja, bakteriomocz, bronchoskopia, cystostomijny, dezalergizacja, diabetyczny*; fotografia: *aerofotografia, bezmatówkowy, bezodblaskowość, bromografia, czarno-białość, drugo-planowy, dwumatrycowy, dwumigawkowy, dwuobrazkowy, fotoamatorski, fotograficzno-artystyczny, fotolityczny, fotoreporterka*; ekonomia: *gospodarczo-racjonalistyczny, hiper-kapitalistyczny, interwencyjno-regulacyjny, inwestycyjno-kredytowy, kapitalistyczno-gospodarczy, korporacjonizm, liberalistyczno-kapitalistyczny, małopłatny, małorządność, nadwłasność, neomerkantylny, obywatel-ekonomista, oportunistyczno-empiryczny, pozapaństwowy, probabilistyczny, robotniczo-emerytalny*; architektura: *agro-*

-miasto, dezurbanizacja, dwukondygnacyjny, gospodarczo-budowlany, kilkumieszkańcowy, kilkunasto-piętrowy, międzynawowy, międzypiętrowy, modernistycznie, nowozaprojektowany, ogólno-architektoniczny, praurbanistyka, przeciw-konstrukcja, pseudoelewacja itd.),

- b) morfologiczne (*-fobia: anglofobia, chromofobia, fotofobia, frankofobia, germanofobia, heliofobia, hydrofobia, ksenofobia, neofobia, polonofobia, pragmatofobia, programofobia, prusofobia, rusofobia, syfilidofobia, ścisłofobia* itp.),
- c) nazwowe (*Auto-Radio, Auto-Radioskład, Centro-bank, Centro-Cement, Centro-Ćwiek, Centro-Hurt, Centro-Kajet, Centro-Kopyto, Centro-mikro, Centro-Obcas, Centro-papier, Centro-parch, Centro-przewód, Centro-sport, Centro-tutka, Centro-wapno, Foto-Radio, Foto-Radio-Palace, Radioantena, Radio-Audion, Radio-Centrala, Radio-Centrum, Radio-City, Radio-Dynamika, Radiodziennik, Radioekspert, Radioelektor, Radio-Express, Radiofil, Radio-front, Radioglob, Radio-Glob, Radiogram, Radioklim, Radio-Korona, Radiokrążek, Radiokrytyka, Radio-Maszyna, Radio-Milion, Radiomotor, Radio-Motor, Radio-pogotowie, Radio-Rower-Elektro, Radio-Selekt, Radio-skład, Radiosport, Radio-Sport, Radio-Stadion, Super-Radio* itp.),
- d) itp.

Uzyskano unikalny zasób. Wedle wiedzy piszącego te słowa zasób tego typu powstał nigdy wcześniej w takiej skali (kilkadziesiąt tysięcy jednostek). Całość słownictwa może z kolei posłużyć do wielu analiz językoznawczych (morfologicznych, etymologicznych itp.), historycznych, antropologicznych, socjologicznych itd. Materiał uzyskany w analizie wydaje się na tyle cenny poznawczo, że w przeszłości planuje się jego wykorzystanie badawcze.

4. Struktura cytatu fotoleksykograficznego

Komponentami prezentacji wyniku ekscerpcyjnego są:

- a) zlematyzowane hasło (*dictionary entry*),
- b) poświadczenie dokumentacyjne (cytat),
- c) poświadczenie lokalizacyjne (tytuł druku),
- d) poświadczenie chronologiczne (data druku).

Hasło zapisywane jest za pomocą czcionek pogrubionych. Poświadczenie dokumentacyjne (cytat) musi być przedstawione w formie umożliwiającej swobodną lekturę, o ile to możliwe – z uwzględnieniem oryginalnej wielkości druku; tytuł źródła musi być zeskanowany (ukazany w oryginalnej postaci graficznej).

Przykładowa prezentacja wyniku ekscerpcyjnego:

sino-modry

0003

narzędzia. Środkiem szerokiej rzeki, fala
sino-modra wre, pędzi i szumi — a brzegiem
jak srebrna koronka, lśnią poszarpane, lodo-
we bisior.

KURYER
LITERACJO - NAUKOWY

31-go marca 1930 r.

Zadbano, aby w żadnym miejscu publikacji nie nastąpiło rozerwanie zaprojektowanej całości leksykograficznej, tj. obiektu składającego się z hasła oraz 3 obrazów (cytatu, lokalizacji, chronologizacji).

5. Wnioski

W rezultacie prac, obok wyniku zasadniczego, jakim jest uzyskanie dokumentacji ok. 80 000 jednostek z lat 30. XX wieku, pozyskano także pewne informacje o strukturze Słownika warszawskiego, m.in. dotyczące:

- hasel latentnych, np. *adrianopolski*, *akcjonista*, *arsenowodór*, *astenik* (istnieją w Słowniku warszawskim, ale nie na prawach hasel) itd.
- wariantów ortograficznych, np. *ariostyczny* – *arjostyczny*, *dwu-arkuszowy* – *dwuarkuszowy*, *hugenot* – *hugienot*,
- błędów literowych, np. w formie poprawnej **Glagolicki, Przewiercanie** (w Słowniku warszawskim błędnie zapisane) itd.

Całość wyniku ekscerpcyjnego ujęto w 30 tomach:

tom 1	abchazki	auto-pogotowie
tom 2	auto-polewaczka	centryfugowanie
tom 3	centryfugowy	doping
tom 4	dopingowanie	ekstremistyczny
tom 5	ekstremizm	fotorotacyjny
tom 6	Foto-Salon	hiperfiskalny
tom 7	hiperfollikularny	jednoziarnowy
tom 8	jednozimowy	konstruktywnie
tom 9	konstruktywno-dziejowy	lewicowo-anarchistyczny
tom 10	lewicowo-centrowy	miłoblążeński
tom 11	miłodźwięczny	nicianka

tom 12	niciarnia	oficer-psycholog
tom 13	oficer-psychotechnik	panceltycki
tom 14	pancerka	podsiarkawy
tom 15	podsiarkowy	pół-jesienny
tom 16	półjęczec	prokominternowski
tom 17	pro-komuna	przesprytny
tom 18	przesprzedaz	radio-meteorologiczny
tom 19	radiometr	rudowasy
tom 20	rudowiec	siwopstry
tom 21	siwo-siny	stratosferyczność
tom 22	stratosferyczny	średniobułgarski
tom 23	średniocenny	triola
tom 24	triolet	urzędowo-statystyczny
tom 25	urzędowo-uroczysty	witaminowo-wapniowy
tom 26	witaminowy	wydajniejszy
tom 27	wydalniczy	zespół-pułk
tom 28	zespół-synteza	żyźniejszy
tom 29	reprodukty	reprodukty
tom 30	reprodukty	reprodukty

Tomy 29 i 30 wypełnione zostały reproduktami z lat 30. XX wieku, np. *artykuł pierwszej potrzeby, budownictwo mieszkaniowe, czas to pieniądz, Follow the Fleet, Gabriel Narutowicz, giełda pracy, gwiazda filmowa, hocki klocki, home fleet, inne cele prywatne, isticie po amerykańsku, kajet zdrowia, klasa obiektywna, linia Zygfyryda, literatura piękna, między wierszami, nigdy nie wiadomo, ochrona specjalna, ojciec narodu, ostre pogotowie, pacta conventa, Państwowa Rada Ochrony Przyrody, Powrót posła, prawo podmiotowe, prawo silniejszego, rozdział kościoła od państwa, Rozkoszna dziewczyna, skazany na śmierć, sprawiedliwy gniew, święto lasu, święto wiosny, święty gaj, Taras Bulba, w górę serca, w stanie wojny, wróg narodu* itd.

Materiał ten jest w tej chwili największym dostępnym zbiorem dla badań z okresu lat 30. XX wieku. Może on posłużyć przede wszystkim do dalszych analiz językoznawczych, w tym do budowy wielkich modeli morfologicznych, zwłaszcza dotyczących dynamiki rozwoju słownictwa w pierwszej połowie XX wieku.

Dodać tu należy, że istnieje polska tradycja tworzenia wielkich modeli leksykalno-derywacyjnych, opracowanych w XX i XXI w. W tradycję tę wpisują się takie prace, jak np. Jadacka, 2001; Smółkowa, 2001; Kleszczowa, 2003; Waszakowa, 2005, itp. Cechą tych ujęć jest przyjęcie i analiza materiału empirycznego w wielkości ok. 3000–5000 jednostek leksykalnych. Podstawowym

parametrem, względem którego następuje analiza danych, jest parametr techniki derywacji oraz wykładnika morfologicznego. Tak więc podział i opis rzeczywistości leksykalnej w tych modelach odbywa się według przykładowych zmiennych, np.: derywacja afiksalna a bezafiksalna czy derywacja na *-ka* lub *-owiec*. W ten sposób konstruowane są głównie opracowania słowotwórcze, np. Kleszczowa, 2003; Jadacka, 2001; Waszakowa, 2005. Tzn. w pracach tych dokonuje się podziału zgromadzonego materiału leksykalnego na zbiory, które autorzy na wstępnych etapach prac badawczych uznają za istotne ze względu na zamierzony cel – przebadanie ewolucji systemu słowotwórczego danego języka, np. polszczyzny.

W konsekwencji uzyskujemy opis języka dokonany na podstawie pewnej próby materiałowej (jak wskazano wyżej: w wielkości ok. 3000–5000) oraz ze względu na przyjęte parametry badań. Opis ten najczęściej obejmuje informacje kwantytatywne (głównie dotyczące produktywności danych wykładników morfologicznych) oraz interpretację tych danych (np. w zakresie wyjaśnienia przyczyn zmian morfologicznych). W wypadku modeli Jadacka, 2001 czy Waszakowa, 2005 dochodzi także komponent prognostyczny, tj. teorie Jadackiej i Waszakowej pozwalają na wysunięcie pewnej klasy sądów o charakterze przewidywań dotyczących rozwoju polszczyzny w XXI w.

Wydaje się, że liczba ok. 80 000 wiernie udokumentowanych i schronologizowanych jednostek, jakie dostarcza 30-tomowe *Depozytorium leksykalne języka polskiego*, może stanowić przekonującą podstawę materiałową dla badaczy języka polskiego XX wieku.

Bibliografia

- Górny, M. i P. Wierzchoń. 2010. *Polish digital libraries as a philologist's tool. Based on 666 adjectives from the Digital Library of Wielkopolska*. Poznań: Instytut Językoznawstwa UAM.
- Jadacka, H. 2001. *System słowotwórczy polszczyzny (1945–2000)*. Warszawa: Wydawnictwo Naukowe PWN.
- Kleszczowa, K. 2003. *Staropolskie derywaty przymiotnikowe i ich perspektywiczna ewolucja*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Kotarbiński, T. 1938. O istocie i zadaniach metodologii ogólnej. *Przegląd Filozoficzny* 4. 69–70.
- Matelski, D. 2007. *Polityka Niemiec wobec polskich dóbr kultury w XX wieku*. Toruń: Wydawnictwo Adam Marszałek.
- Mazurek, C., Stroiński, M., Werla, M. i J. Węglarz (red.). 2009. *Polskie biblioteki cyfrowe 2008*. Poznań: Ośrodek Wydawnictw Naukowych PAN.
- Mazurek, C., Stroiński, M., Werla, M. i J. Węglarz. 2009. „Infrastruktura bibliotek cyfrowych w sieci PIONIER”. W zbiorze: Mazurek, C., Stroiński M., Werla, M. i J. Węglarz (red.). 9–13.
- Smólkowa, T. 2001. *Neologizmy we współczesnej leksyce*. Kraków: Lexis.
- Waszakowa, K. 2005. *Przejawy internacjonalizacji w słowotwórstwie współczesnej polszczyzny*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Wawrzyńczyk, J. 1999. *Nowe słownictwo polskie. Fikcje i fakty*. Warszawa: Instytut Informacji Naukowej i Studiów Bibliologicznych.

- Wawrzyńczyk, J. 2000. *Słownik bibliograficzny języka polskiego. Wersja przedelektroniczna. Tom 1*. Warszawa: Instytut Informacji Naukowej i Studiów Bibliologicznych.
- Wawrzyńczyk, J. 2009. *Autosuplement do Słownika warszawskiego*. Poznań: Sorus.
- Wawrzyńczyk, J. (red.). 2009. *Czterdzieści lat minęło... nad „Słownikiem Doroszewskiego”*. Warszawa: TAKT.
- Wawrzyńczyk, J. 2011. *Słownictwo nowopolskie. Redatacje*. Warszawa: Bel Studio.
- Węglarz, J. 1981. *Sterowanie w systemach typu kompleks operacji*. Warszawa: PWN.
- Wierchoń, P. 2007. W poszukiwaniu czasowników nieznanych lingwist(k)om. *Język. Komunikacja. Informacja* 2. 149–197.
- Wierchoń, P. 2008a. *Fotodokumentacja. Chronologizacja. Emendacja. Teoria i praktyka weryfikacji materiału leksykalnego w badaniach lingwistycznych*. Poznań: Instytut Językoznawstwa UAM.
- Wierchoń, P. 2008b. *Jaskółki przejawów internacjonalizacji w słowotwórstwie współczesnej polszczyzny w materiałach z lat 1894–1984. Tylko sto przykładów*. Łask: LEKSEM.
- Wierchoń, P. 2008c. *ANTI*. Poznań: Instytut Językoznawstwa UAM.
- Wierchoń, P. 2008d. *Kotuś. «Verba polona abscondita...» (w fotodokumentacji). Szkic lingwochronologizacyjny. Centuria pierwsza*. Poznań: Instytut Językoznawstwa UAM.
- Wierchoń, P. 2009a. *Dlaczego fotodokumentacja? Dlaczego chronologizacja? Dlaczego emendacja? Instalacja gazowa, parking podziemny i „odległość niezerowa”*. Poznań: Instytut Językoznawstwa UAM.
- Wierchoń, P. 2009b. „Fotodokumentacja 3.0”. *Język. Komunikacja. Informacja* 4. 63–80.
- Wierchoń, P. 2010a. „Torując drogę teorii lingwochronologizacji”. *Investigationes Linguisticae* XX. 105–186.
- Wierchoń, P. 2010b. „Lingwochronografia na usługach słowotwórstwa gniazdowego”. *Kwartalnik Językoznawczy* 1.2. 50–64.
- Wierchoń, P. 2010c. *Depozytorium leksykalne języka polskiego. Nowe fotomateriały z lat 1901–2010. Tom I*. Warszawa: Bel Studio.
- Wierchoń, P. 2011a. *Depozytorium leksykalne języka polskiego. Nowe fotomateriały z lat 1901–2010. Tom II*. Warszawa: Bel Studio.
- Wierchoń, P. 2011b. *Depozytorium leksykalne języka polskiego. Nowe fotomateriały z lat 1901–2010. Tom III*. Warszawa: Bel Studio.
- Wierchoń, P. 2014. *Depozytorium leksykalne języka polskiego. Fotosuplement do Słownika warszawskiego. Tom XI–XL*. Warszawa: Bel Studio.