

Uwagi wstępne

Projekt o nazwie „Narodowy Fotokorpus Języka Polskiego” (dalej: NFJP) realizowany był od połowy roku 2014 do połowy roku 2020 dzięki wsparciu finansowemu udzielonemu w formie grantu przyznanego przez Ministra Nauki i Szkolnictwa Wyższego w ramach rządowego konkursu Narodowy Program Rozwoju Humanistyki. W stosownej umowie, zawartej między Ministerstwem a Uniwersytetem Warszawskim jako jednostką naukową prowadzącą ów grant, znajduje się informacja o podtytule projektu: „Fotodokumentacja słownictwa XX w.”.

Ogólnie, w celach marketingu naukowego przedsięwzięcie było już przedstawiane w kilku niewielkich publikacjach (por. Wawrzyńczyk, Wierchoń 2016; Wawrzyńczyk 2017; Wawrzyńczyk, Wierchoń 2017). Ma ono charakter unikatowy, jego celem była konstrukcja największego w dziejach leksykografii języka polskiego zbioru danych – poświadczeń cytatowych (głównie zdaniowych, nierzadko wielozdaniowych) dla około 250 000 haseł słownikowych, z jednego tylko stulecia, poświadczeń jednostek ujmowanych jako obiekty leksykograficzne zgodnie z kilkunastowieczną tradycją, której istotnym składnikiem, jednym z fundamentów, jest 11-tomowy *Słownik języka polskiego*, opracowany przez zespół pod kierunkiem Witolda Doroszewskiego (por. Doroszewski 1958–1969). Słownik ten, dziś traktowany powszechnie jako dzieło klasyczne, kapitalne, liczy około 125 000 haseł.

Po sześciu latach pracy nad NFJP nadszedł czas na pewne etapowe podsumowanie naszych wysiłków i zarazem nakreślenie nowych zadań, nowych perspektyw, by uzyskane dotychczas wyniki można było implementować owocnie (i równocześnie autokrytycznie) w przyszłości.

Pryncypialną nowość w opracowaniu naszego ćwierćmilionowego zbioru haseł polszczyzny XX wieku stanowi forma podawcza: poświadczenia cytatowe prezentowane w NFJP, a umieszczone w najdostępniejszej bibliotece świata, czyli w internecie (na stronie: www.nfjp.pl). To nie ręcznie ekscerpowane wyimki z tekstów źródłowych, lecz fotocyfry, gwarantujące absolutną zgodność wyimka z jego źródłem drukowanym (ewentualnie, zupełnie wyjątkowo – rękopiśmiennym). Brak tej zgodności był i nadal bywa – w nowych słownikach, nietworzonych z zastosowaniem techniki fotoekscerpcji –

przyczyną najrozmaitszych licznych błędów, zniekształceń, w tym powoływania do życia, wprowadzania do obiegu fikcyjnych bytów leksykograficznych, które w anglosaskiej literaturze krytycznej określa się mianem *ghost words* (por. w odniesieniu do opracowania: Doroszewski 1958–1969 pracę Wawrzyńczyk 2010 i Wawrzyńczyk 2016).

Zasada doboru haseł z ustalonego zbioru tekstów źródłowych do hasłownika NFJP – w jego części dwudziestowiecznej – jest elementarnie prosta: trafić do niego powinna każda jednostka nieobecna w dziele Władysława Kokowskiego, pierwszym polskim słowniku ortograficznym XX wieku (Kokowski 1903). W efekcie na podstawie kryterialnej bazy Władysława Kokowskiego sprzed niemal 120 lat powstała baza leksykalna o znacznej objętości, wykraczająca bardzo mocno, zarówno pod względem ilościowym, jak i jakościowym, poza którykolwiek słownik ogólny (czy to tak zwane narodowe, czy inne zasłużone dzieła) języka polskiego.

Cel projektu

Uzyskana baza ma – dzięki nadzwyczaj poważnej liczbie innowacyjnych w formie poświadczeń fotocytatowych – określone znaczenie dla badań nad zasobem i rozwojem słownictwa polszczyzny XX wieku. Każdy ekscerpt jest dokładnie zlokalizowany, to znaczy podaje się nazwę dokumentu, z którego pochodzi. Jednocześnie, obligatoryjnie, ekscerpty zostały dokładnie scharakteryzowane pod względem chronologicznym, czyli odnotowuje się datę roczną ich wystąpienia w tekście.

Zaznaczyć należy, że dla pierwszej połowy XX wieku, w tym dla szczególnie doniosłego w aspekcie dziejo- i kulturotwórczym dwudziestolecia międzywojennego, rejestry leksyki są nader skąpe i ograniczają się do systematycznych ekscerpcji realizowanych przez zaledwie kilku badaczy, konstruktorów NFJP właśnie. Początki rejestracji regularnych sięgają epoki przedfocycytatograficznej, wyznaczają je dwie prace: Wawrzyńczyk 1987 i Wawrzyńczyk 1989. Skokowe zwielokrotnienie danych dwudziestowiecznych otwiera możliwość różnych zastosowań, prowadzenia badań morfologicznych, badań w zakresie kontaktów międzyjęzykowych na poziomie leksykalnym, badań w zakresie frazeologii (w najszerszym rozumieniu tego dziś już nieostrego terminu), analiz stylometrycznych itp.

Podstawa metodologiczna

W zakresie idei inicjującej masową ekscerpcję słownictwa XX wieku najwięcej waży prace dokumentacyjne Jana Wawrzyńczyka (por. Bibliografia), które zaczęły się pojawiać regularnie od końca lat 80. XX wieku. Istota tych prac polega na sięgnięciu do materiałów tekstowych, które (1) nigdy wcześniej nie były ekscerpowane pod kątem leksykograficznym bądź (2) ekscerpowane, wybierane do ekscerpcji były, ale wyrywkowo, bez jakiegokolwiek planu, niesystematycznie.