

# Jerzy Woronczak

---

## Z warsztatu filologa-statystyka

---

Biuletyn Polonistyczny 5/14, 9-11

---

1962

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

5. Zespół leksykograficzny II we Wrocławiu, Rynek 9-11.

Kierownik: prof. Stanisław Bąk.

Pracownicy: Andrzej Cieński, Teresa Kaufmanowa, Janina Pereko-  
wa, Lucyna Woronczakowa.

#### P r z y p i s y :

- 1 Poprzednia informacja o "Słowniku" została umieszczona w 2 zeszytach Biuletynu (czerwiec 1958), ss. 5-6, 11-14.
- 2 "Słownik polszczyzny XVI wieku". Zeszyt próbny. Wrocław 1956.
- 3 "Słownik polszczyzny XVI wieku". Uwagi i recenzje. Warszawa 1957.

#### Z WARSZTATU FILOLOGA-STATYSTYKA

Metodami statystycznymi zainteresowałem się w przekonaniu, że zastosowanie ich w pracach literaturoznawczych pozwala, z jednej strony, na weryfikację stawianych twierdzeń, z drugiej zaś, na rozszerzenie zakresu badań na problemy nie poruszane z racji braku odpowiednich narzędzi badawczych.

Pierwsze moje prace wykonane przy użyciu techniki statystycznej dotyczyły wersyfikacji. W artykule "Z badań nad wierszem Biernata z Lublina" ("Pamiętnik Literacki", r. 49 (1958), z.3, s. 97 - 118) przeprowadziłem statystyczny dowód twierdzenia Brücknera, słusznego, ale nie popartego argumentami, że pierwotna, nie znana nam obecnie redakcja "Ezopa" Biernata z Lublina podległa w XVI w. unowocześniającym przeróbkom językowym, od których ucierpiała m.in. sylabiczna regularność wiersza. Równocześnie jednak dowodziłem, że część odchyłeń od sylabizmu ścisłego trzeba policzyć na karb samego autora. Podałem też charakterystykę ilościową kilku innych cech wiersza i języka Biernata (rytmika klauzuli, rym, szyk przydawki, ściągnięte i nie ściągnięte formy zaimków dzierżawczych), częściowo w porównaniu z innymi utworami XVI w. i omówiłem przykładowo technikę obliczania niektórych zastosowanych wskaźników statystycznych.

Na danych statystycznych był również oparty mój referat "Elementy średniowieczne w wersyfikacji polskiej XVI wieku i ich przemiany" ("Z polskich studiów slawistycznych". Prace

historycznoliterackie na IV Międzynarodowy Kongres Słowistów w Moskwie 1958, Warszawa 1958, s. 243 - 269). Zgodnie z charakterem pracy pominąłem w niej zupełnie "kuchnię statystyczną", a większość koniecznych danych liczbowych umieściłem w osobnej tabeli.

Pominięty w poprzedniej pracy zagadnieniom statystycznego warsztatu poświęciłem referat wygłoszony na warszawskiej konferencji poetyki w 1960 r. pt. "Statistische Methoden in der Verslehre" (Poetics, Poetyka, Poetika, Warszawa, Mouton-PWN 1962). Omówiłem w nim przykładowo zastosowanie teorii serii do badania kompozycji wersyfikacyjnej utworów napisanych wierszem wolnym oraz metody analizy rozkładu długości wersów różnych typów słowiańskiego wiersza asylabicznego, a w związku z nim podałem również charakterystykę rozkładu długości polskiego zdania w prozie. Poszerzona, dotąd nie opublikowana forma tego referatu omawia jeszcze problemy statystyczne związane z automatyczną sylabotoniacją wiersza sylabicznego, rytmiką klauzuli wierszowej i określeniem bogactwa słownikowego rymów.

Ten ostatni problem stanowił dla mnie punkt wyjścia do zajęcia się statystyką słownictwa już nie samych rymów, ale całych utworów, szczególnie zaś możliwością syntetycznej charakterystyki zasobu leksykalnego tekstów pod względem jego bogactwa i rozkładu częstości wyrazów. W tym celu należy obliczyć wskaźniki, które by były niezależne od długości badanego utworu, a przy jednorodności opracowywanego tekstu równe (oczywiście w granicach dopuszczalnych odchyień) tak dla całego utworu, jak i dla jego dowolnie długich części. Niektóre spośród zaproponowanych w dotychczasowej literaturze wskaźników nie odpowiadają tym warunkom, gdyż zmieniają się w zależności od rozmiaru tekstu, nie mogą więc służyć do porównywania utworów o różnej długości. Inne, pozbawione tej wady, nie dają wyczerpującej charakterystyki, a dla, moim zdaniem, najlepszych nie opublikował projektodawca metody obliczenia. W tej sytuacji, przyjmąwszy za podstawę wskaźniki B. Mandelbrota, opracowałem sposób ich obliczenia i zaprojektowałem graficzną metodę przedstawiania uzyskanych wyników. Polega ona na tym, że utworowi przyporządkowuje się punkt na specjalnej siatce współrzędnych, którego położenie zależne jest od obliczonych wskaź-

ników. Rachunki, jakich dokonałem dla 9 utworów (w większości polskich tekstów XVI w.), dały zadawalające wyniki. Rezultaty badań przedstawiłem latem 1961 r. na konferencji poetyki w Jabłonnie w referacie "Metody obliczania wskaźników bogactwa słownikowego tekstów".

Przy obliczaniu wskaźników Mandelbrota wychodzi się od listy frekwencyjnej jednostek leksykalnych tekstu. Sporządzenie jej jest pracą bardzo żmudną i dlatego pożądane byłoby oparcie się nie na całym tekście utworu, lecz tylko na jego próbkę. Obecnie pracuję nad określeniem wielkości próbki, wystarczającej dla oszacowania wskaźników bogactwa słownikowego tekstu z żadaną dokładnością. Dotychczasowe wyniki pozwalają na stwierdzenie, że wskaźniki obliczone na podstawie zbyt małych próbek ciągłych (do około 1000 wyrazów) wykazują w stosunku do wskaźników obliczonych dla całego utworu nie tylko odchylenia przypadkowe, ale także różnicę systematyczną, wynikającą z tego, że ich bogactwo słownikowe jest z reguły nieco mniejsze niż bogactwo słownikowe całego utworu. Stąd wystarczająca próbka ciągła musi mieć długość rzędu kilku tysięcy wyrazów. Próbki rozproszone, szczególnie jeśli będziemy losować do nich luźne wyrazy, nie wykazują wspomnianej różnicy systematycznej. Mogłyby się one zamykać w granicach 500 - 1000 wyrazów. Trzeba jednak zaznaczyć, że luźne wyrazy można wygodnie pobierać do próbki jedynie z tekstów prozaicznych. Próbki z utworów wierszowanych należy raczej budować z pełnych wersów, a to ze względu na zasadnicze różnice między wyrazami stojącymi na początku, w środku i na końcu wersów.

Dr Jerzy Woronczak  
Instytut Badań Literackich

Z PRAC NAD WYDAWNICTWEM "OBRAZ LITERATURY POLSKIEJ  
XIX i XX WIEKU"

Wydawnictwo, którego tytuł roboczy brzmiał "Dwieście lat literatury polskiej", zacznie niebawem ukazywać się drukiem nakładem "Wiedzy Powszechnej" pt. "Obraz literatury polskiej XIX i XX wieku".