

# **Podręcznik do korpusologii. Rafał L. Górski, Magdalena Król, Maciej Eder, *Zmiana w języku. Studia kwantytatywno-korpusowe*. Kraków: Instytut Języka Polskiego PAN, 2019**

**Maciej Rak**

Uniwersytet Jagielloński w Krakowie  
maciej.rak@uj.edu.pl

Książka trojga autorów – Rafała L. Górskiego, Magdaleny Król i Macieja Edera, *Zmiana w języku. Studia kwantytatywno-korpusowe*, powstała jako wynik prac, które w latach 2014–2018 prowadzono w ramach projektu kierowanego przez R.L. Górskiego i finansowanego przez Narodowe Centrum Nauki – *Przebiegi zmian gramatycznych i leksykalnych w historii języka polskiego – metody korpusowe i kwantytatywne w językoznawstwie diachronicznym*. Monografia ma zagwarantować lepszą rozpoznawalność projektu w publikacjach, które powołują się na jego wyniki. Taka praktyka jest już stosowana, np. w odniesieniu do *Wielkiego słownika języka polskiego PAN* (por.: Żmigrodzki et al. 2018) i *Narodowego Korpusu Języka Polskiego* (por.: Bańko et al. 2012).

Książka, o której mowa, nie jest obszerna (liczy 153 strony), co należy postrzegać raczej jako jej zaletę. Składa się z 5 rozdziałów i zakończenia. Bibliografia (s. 143–149) w dużej mierze jest obcojęzyczna (dokładniej: anglojęzyczna), pośrednio dowodzi więc, że badania kwantytatywno-korpusowe są rozwijane zwłaszcza w odniesieniu do angielszczyzny.

Rozdział pierwszy – zresztą jak cała książka – przekonuje, że w XXI wieku w językoznawstwie będą dominowały badania oparte na korpusach tekstów i metodach ilościowych. Ich zapowiedzią są zwłaszcza prace dotyczące historycznego językoznawstwa anglistycznego. Z tego kręgu naukowego idą bowiem inspiracje i tutaj mamy prekursorskie prace, co można wyjaśnić nie tylko niekwestionowaną obecnie rangą angielszczyzny, lecz także dość dawną (zwłaszcza w porównaniu z takimi językami, jak polszczyzna) tradycją piśmienniczą i wieloma zachowanymi dokumentami.

W kolejnym rozdziale autorzy dokładnie (by nie rzec, krok po kroku) zdają sprawozdanie z tego, jak tworzyli korpus i pracowali na nim. Poruszone są tu m.in. sprawy konieczności ujednolicenia ortografii i odpowiedniej anotacji morfosyntaktycznej leksemów. Jasno z tego

wynika, że na powstawanie korpusu składa się szereg działań, w których musi uczestniczyć odpowiednio przygotowany badacz. Już to przekonuje, że np. przed dialektologami, zamierzającymi pracować na korpusach gwarowych, droga do tego typu badań jest znacznie dłuższa i bardziej utrudniona. W praktyce trzeba bowiem zdobyć teksty, czyli nagrać je u odpowiednich informatorów, a później przepisać (na razie programy do syntezy pisma na podstawie mowy nie sprawdzają się w wypadku gwar) w odpowiedniej, jak sądzę, półfonetycznej ortografii (by korpusy z różnych regionów mogły się uzupełniać, ten rodzaj ortografii powinien być wystandaryzowany, co oczywiście wcale nie jest proste w odniesieniu do całego polskiego obszaru językowego). Z wszystkimi tego typu rozterkami i problemami mierzyli się wykonawcy projektu *Korpus Spiski* (Grochola-Szczepanek et al. 2019). Wróćmy jednak do książki R.L. Górskiego, M. Król i M. Edera. Rzecz jasna, autorom nie chodzi tu *stricto* o sam korpus, lecz o metody (które mogą znaleźć zastosowanie przy opisie różnego materiału empirycznego, także np. gwarowego) oraz o „rozpowszechnienie idei otwartego oprogramowania i udostępniania danych użytych w eksperymentach” (s. 10). Dowiadujemy się więc, że w zależności od konkretnego zapotrzebowania badawczego korpus może być modyfikowany, rozszerzany o dodatkowe podkorpusy. Jeśli zaś idzie o oprogramowanie, wszystkie omówione w książce wyniki osiągnięto przy wykorzystaniu bezpłatnych programów statystycznych i korpusowych (głównie AntConc i język programowania R). Co więcej, dane wejściowe zostały zdeponowane w ogólnodostępnym repozytorium internetowym GitHub, a to daje możliwość zweryfikowania wyników ogłoszonych w recenzowanej pracy. Nie jest to częsta sytuacja w polskiej humanistyce. Zwykle otrzymujemy wnioski, ale nie przychodzi nam do głowy pomysł powtórzenia wykonanego już eksperymentu. Rozwiązanie wdrożone przez Górskiego, Król i Edera stanowi zachętę dla czytelnika i może być świetnym ćwiczeniem pokazującym, jak pracować z korpusami tekstów.

W trzecim rozdziale pt. *Dynamika zmian językowych* omówiono – obrazując to odpowiednimi rysunkami – przebieg zmian (albo też konkurencję) w obrębie leksemów: *więtszy* > *większy*, *barzo* > *bardzo*, *inszy* > *inny*, *wszytek* > *wszystek*, flektywów: *-bych* > *-bym*, *-bychmy* > *-byśmy*, oraz grupy *-ir-* > *-er-*. Z tych analiz jasno wynika, że zastępowanie formy recesywnej formą innowacyjną nie zawsze przebiega w sposób bardzo regularny, zgodnie z modelem regresji logistycznej, bowiem zmiana językowa jako proces *de facto* społeczny jest zależna od różnych zakłócających ją czynników. Oczywiście jest to proces, którego dynamika jest zmienna zwłaszcza w zależności od grupy społecznej i obszaru (czynnik geograficzny jest więc nie mniej ważny niż społeczny). Zjawiska omówione w tym podrozdziale zaciekawiają czytelnika także z tego względu, że mówią o dynamice zmian językowych, pokazują różne tempo procesów, a

także ich współbieżność. Szczególnie ciekawie pod tym względem mogłyby się przedstawiać rysunki, które jednak są nieczytelne. Brak kolorów (to samo dotyczy wykresów kołowych) sprawia, że nie daje się z nich w odpowiedni sposób skorzystać.

*Klasyfikacja maszynowa w językoznawstwie diachronicznym* to tytuł kolejnego rozdziału. Głównym jego zamysłem jest wskazanie momentu największej dynamiki zmian w dziejach języka, co stanowi podstawę wszystkich ze znanych nam periodyzacji polszczyzny. Do tej pory wyznaczanie granic epok dokonywało się w pewnym stopniu intuicyjnie, niebagatelną rolę odgrywał tu także autorytet badacza – autora periodyzacji. Propozycje zawarte w tym rozdziale są niezwykle nośne intelektualnie i przekonujące, jak choćby uwagi na temat przyimków w historii polszczyzny. Niestety i tu wykresy są nieczytelne.

Rozdział piąty autorzy w całości poświęcili frekwencji imiesłowu przysłówkowego uprzedniego. Co może zaskakiwać, zwłaszcza z perspektywy gramatyki opisowej języka polskiego, formy na *-wszy*, *-wszy* obecnie (w polszczyźnie XX w. i XXI w.) odzyskują pozycję, zwiększając frekwencję. W szczegółach wygląda to następująco: najpierw były to formy bardzo rzadkie, ale zyskały bardzo dużą popularność w XVII w., później frekwencja znaczenie się zmniejszyła, by osiągnąć najniższy poziom na przełomie XVIII i XIX w. Dziś znów obserwujemy wzrost frekwencji – oczywiście uwagi te dotyczą (podobnie jak wszystkich innych zjawisk omówionych w tej książce) polszczyzny drukowanej. Zmiany we frekwencji imiesłowu przysłówkowego uprzedniego można wyjaśnić – jak podają autorzy – na dwa sposoby: albo jest to moda językowa, albo zmieniła się produktywność tej formy gramatycznej (zmniejszyły się ograniczenia stawiane czasownikom, od których jest tworzony imiesłów uprzedni).

Książkę zamyka syntetyczne podsumowanie. Po raz kolejny podkreślam lapidarność stylu, którą traktuję także jako wielki atut omawianej monografii. Po przybliżeniu jej zawartości wyjaśnię tytuł niniejszej recenzji.

W *Wielkim słowniku języka polskiego* pod red. Piotra Źmigrodzkiego (WSJP PAN) rzeczownik *podręcznik* został objaśniony jako: ‘książka zawierająca zbiór najważniejszych informacji z określonej dziedziny i służąca do nauki’<sup>1</sup>. Można by tę definicję uzupełnić o *jasność* i *przejrzystość*, choć te dwie cechy są w pewnym stopniu implikowane przez stwierdzenie *służąca do nauki*.

Ponieważ moja styczność z korpusologią jest niewielka, *Zmianę w języku* traktuję jako najlepszy podręcznik do tego, jak tworzyć korpusy i na ich podstawie prowadzić badania.

<sup>1</sup> [https://wsjp.pl/index.php?id\\_hasla=61867&ind=0&w\\_szukaj=podr%C4%99cznik](https://wsjp.pl/index.php?id_hasla=61867&ind=0&w_szukaj=podr%C4%99cznik) (dostęp: 20 IX 2021).

Książka R.L. Górskiego, M. Król i M. Edera zawiera zbiór najważniejszych informacji na temat badań kwantytatywno-korpusowych, przekazanych w sposób bardzo jasny i przejrzysty, w efekcie czego chyba żaden czytelnik nie poczuje się tą lekturą przytłoczony. Struktura książki, jej cel, styl i możliwość samodzielnego powtórzenia omówionych w niej badań sprawiają, że *Zmiana w języku* powinna stać się uniwersyteckim podręcznikiem do korpusologii.

### **Bibliografia**

- Bańko, Mirosław et al. (red.) (2012) *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Grochola-Szczepanek, Helena et al. (2019) „Korpus języka mówionego mieszkańców Spisza”, *LingVaria* XIV, nr 1(27); 165–180.
- WSJP PAN: Żmigrodzki, Piotr (red.) *Wielki słownik języka polskiego PAN*. <http://wsjp.pl>.
- Żmigrodzki, Piotr et al. (red.) (2018) *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*. Kraków: Libron.