

Annales Universitatis Paedagogicae Cracoviensis

Studia Linguistica 19 (2024)

ISSN 2083-1765

DOI 10.24917/20831765.19.13

Katarzyna Ostrowska

Uniwersytet Jana Kochanowskiego, Kielce, Polska

ORCID 0000-0001-9638-0308

Marek Ruszkowski

Uniwersytet Jana Kochanowskiego, Kielce, Polska

ORCID 0000-0001-9305-6043

Wskaźnik ilościowego bogactwa słownictwa jako element leksykalnej charakterystyki współczesnego polskiego reportażu książkowego

Słowa kluczowe: bogactwo słownictwa, frekwencja leksemów, frekwencja słowoform, reportaż książkowy

Keywords: vocabulary richness frequency of lexemes, frequency of word forms, book reportage

Wprowadzenie

Wskaźniki ilościowe od kilkudziesięciu lat są coraz częściej stosowane w językowej charakterystyce tekstu (np. Guiraud 1966; Pisarek 1969, 1972; Sambor 1972; Sigurd 1975: 157–159; Zembaty-Michalakowa 1982; Zarębina 1985; Rachwałowa 1986: 22–70; Zgółkowa 1987; Hammerl, Sambor, 1990; Łojek 2001; Pawłowski 2001; Ruszkowski 2004a, 2004b, 2022; Ostrowska 2023a; Ostrowska, Ruszkowski 2024: 38–48). Pozwalają one w miarę obiektywnie przeanalizować te cechy stylistyczne tekstu, które bez zastosowania wskaźników podlegałyby tylko ocenie intuicji badacza, a to z natury rzeczy jest podejściem subiektywnym. Jak pisała Maria Mayenowa (1965: 14), „intuicją statystyczną posługują się nawet ci, którzy w teorii są jej gorącymi przeciwnikami”.

Jednym z najważniejszych parametrów pozwalających scharakteryzować leksykę tekstu jest **wskaźnik ilościowego bogactwa słownictwa** (I_{bog}). W badaniach stylistycznych stosowane są trzy wskaźniki, dzięki którym można precyzyjnie określić bogactwo leksyki (Sambor 1972: 219–220).

Pierwszy z nich opracował francuski lingwista Pierre Guiraud (1954: 52):

$$I_{\text{bog}} = \frac{V}{\sqrt{2N}}$$

Autorem drugiego jest Władysław Kuraszkiwicz (1958: 240):

$$I_{\text{bog}} = \frac{V}{\sqrt{N}}$$

Trzeci opracował słowacki językoznawca Jozef Mistrik (1967: 45):

$$I_{\text{bog}} = \frac{20V}{N}$$

V – liczba różnych leksemów w tekście, czyli słownik tekstu

N – długość tekstu, czyli liczba wszystkich słowoform

Wskaźniki te różnią się jedynie liczbowo, „mają natomiast identyczną interpretację” (Sambor 1972: 220). Jednak najlepszy, najwygodniejszy podczas dokonywania obliczeń i zapewne stosowany najczęściej jest wskaźnik J. Mistrika, który zostanie wykorzystany w niniejszym artykule.

Należy jednak pamiętać, że wskaźnik I_{bog} jest porównywalny tylko wtedy, gdy badamy teksty lub statystycznie reprezentatywne próbki tekstów o jednakowej długości. Dzieje się tak, ponieważ w miarę wzrostu długości tekstu (N) przyrost słownictwa (V) jest coraz mniejszy. Przykładem mogą być reportaże poddane analizie językowej. Najobszerniejszy z nich (D. Rosiak, *Ziarno i krew. Podróż śladami blisko-wschodnich chrześcijan*) zawiera łącznie 100 433 słowoformy i 10 214 leksemów, czyli wartość badanego wskaźnika wynosi 2,0. Z kolei najkrótszy reportaż (J. Mikołajewski, *Wielki przypływ*) liczy N = 20 438 i V = 4203, czyli $I_{\text{bog}} \approx 4,1$. Na pozór więc tekst J. Mikołajewskiego jest ponad dwa razy bogatszy leksykalnie niż reportaż D. Rosiaka, co nie jest prawdą. *Ziarno i krew* jest tekstem prawie pięć razy dłuższym niż *Wielki przypływ*, przez co przyrost słownictwa (V) po przekroczeniu pewnej długości (N) znacznie maleje.

Podobne przykłady można przytoczyć, przywołując teksty prozy artystycznej. James Joyce w obszernej powieści *Ulisses* użył 260 430 słowoform (N) i 29 899 leksemów (V) (Zipf 1935; za Wierzbicka 1987: 144). Tak więc $I_{\text{bog}} = \frac{20 \cdot 29899}{260430} \approx 2,3$. Z kolei *Lalka* Bolesława Prusa zawiera 235 718 wyrazów i 14 382 leksemy (Smółkowska 1974: 72, 87), czyli wskaźnik ilościowego bogactwa słownictwa wynosi 1,3 ($I_{\text{bog}} = \frac{20 \cdot 14382}{225718} \approx 1,3$). Obie wartości są bardzo niskie, gdyż badano całą długość (N) tekstów, które liczą po dwieście kilkadziesiąt tysięcy słowoform. Nie można ich porównywać z innymi danymi, które oparte są na znacznie mniejszych wielkościach (N). Dlatego też – aby porównywać teksty o różnej (lub podobnej) stylistycznej proveniencji – najczęściej przyjmuje się wartość 10 000 słowoform (N) i bada, jaka wartość (V) mieści się w tej próbie (np. Laskowska 1985: 152; Zarębina 1985: 12; Rachwałowa 1986: 23–24; Gąsiorek 1997: 10; Łojek 2001: 36; Vaškevič-Buš 2017: 49). Na przykład w języku dzieci w wieku szkolnym (szkoła podstawowa) I_{bog} obliczony na podstawie

$N = 10\ 000$ wynosi 3,7 (Zarębina 1985: 130), a więc na pozór jest znacznie większy niż w powieściach J. Joyce'a i B. Prusa. W rzeczywistości jest dokładnie odwrotnie, ale podanie konkretnych danych liczbowych wymagałoby ujednoczenia wielkości próby (N).

Założenia artykułu

Przedmiotem analizy i interpretacji w tym artykule jest 18 polskich reportaży książkowych z lat 2004–2019 wyróżnionych Nagrodą im. Beaty Pawlak: *Jadąc do Babadag* Andrzeja Stasiuka (2004), *Japoński wachlarz* Joanny Bator (2004), *Izrael już nie frunie* Pawła Smoleńskiego (2006), *Blondynka na Kubie. Na tropach prawdy i Ernesta Che Guevary* Beaty Pawlikowskiej (2006), *Gottland* Mariusza Szczygła (2006), *Dym się rozwiera* Jacka Milewskiego (2008), *Oko świata. Od Konstantynopola do Stambułu* Maxa Cegielskiego (2009), *Czwarty pożar Teheranu* Marka Kęskrawca (2010), *Zabójca z miasta moreli. Reportaże z Turcji* Witolda Szablowskiego (2010), *Abchazja* Wojciecha Góreckiego (2013), *Dom nad rzeką Loes* Mateusza Janiszewskiego (2014), *Dryland* Konrada Piskaty (2014), *Wielki przyptyw* Jarosława Mikołajewskiego (2015), *Ziarno i krew* Dariusza Rosiaka (2015), *Dybuk. Opowieść o nieważności świata* Krzysztofa Kopczyńskiego i Anny Sajewicz (2017), *Deutsche nasz. Reportaże berlińskie* Ewy Wanat (2018), *Raban. O kościele nie z tej ziemi* Mirosława Wlekłego (2019), *Wędrowny zakład fotograficzny* Agnieszki Pajączkowskiej (2019) (zob. Ostrowska 2023b: 24).

Z początku każdego z reportaży wyekscerpowano po 10 000 słowoform (N), aby zachować postulat jednakowej wielkości próby, a w rezultacie uzyskać możliwość porównywania wartości wskaźnika. Pozyskane próbki reportaży książkowych poddano analizie ilościowej za pomocą Literackiego Eksploratora Maszynowego¹ z infrastruktury CLARIN-PL, dzięki czemu wyznaczono statystykę leksemów i słowoform w badanych reportażach. Następnie, korzystając ze wzoru J. Mistrika (N – liczba słowoform, czyli po 10 000 z początku każdego reportażu, V – liczba leksemów w każdej z tych prób), obliczono wskaźnik I_{bog} dla poszczególnych tekstów, wahający się między 5,2 (D. Rosiak, *Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan*) a 6,8 (M. Janiszewski, *Dom nad rzeką Loes*), oraz średnią dla reportażu $I_{\text{bog}} \approx 5,9$.

Ostatnim etapem było ustalenie – na tle porównawczym innych typów tekstów – w którym miejscu znajduje się wskaźnik I_{bog} dla współczesnego polskiego reportażu książkowego. W związku z tym wykorzystano badania Marii Zarębiny (1985: 130) dotyczące słownictwa polszczyzny mówionej oraz Marii Rachwałowej (1986: 24) na temat tekstów naukowych. Wielkość N w badanych próbach jest bardzo zbliżona, wynosi bowiem od 9972 do 10 030, czyli wartości I_{bog} są wiarygodne i porównywalne.

¹ Według zespołu CLARIN-PL: „Główne cechy usługi LEM to: lematyzacja, tagowanie części mowy, generowanie niestandardowych form wyrazowych i tekstów lematyzowanych. Planowane funkcje obejmują: grupowanie tekstów, klasyfikację semantyczną opartą na uczeniu maszynowym i wizualizację jego wyników, modelowanie tematu. LEM jest opracowywany przez CLARIN-PL (Politechnika Wrocławska) we współpracy z Instytutem Badań Literackich Polskiej Akademii Nauk” (Maryl, Piasecki, Walkowiak, LEM, CLARIN-PL, <https://ws.clarin-pl.eu/>; dostęp: 11.01.2024).

Wskaźnik ilościowego bogactwa słownictwa reportażu

Na podstawie analizy poniższej tabeli (tabela 1) można zauważyć, że rozpiętość wskaźnika bogactwa ilościowego słownictwa w badanych tekstach jest znaczna, wynosi bowiem 1,6. Najbardziej bogate słownictwo zawiera reportaż M. Janiszewskiego pt. *Dom nad rzeką Loes* ($I_{\text{bog}} = 6,8$), najmniej zaś *Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan* D. Rosiaka ($I_{\text{bog}} = 5,2$). Dla porównania należy dodać, że wartość I_{bog} w polszczyźnie pisanej wynosi 6,6 (Zarębina 1970: 33–46), a w polszczyźnie mówionej – 4,8 (Zarębina 1971: 336–347), jest więc znacząco niższa.

Tabela 1. Wartość wskaźnika ilościowego bogactwa słownictwa we współczesnych polskich reportażach książkowych.

Lp.	Tytuł reportażu	Frekwencja leksemów (V)	Frekwencja słoform (N)	Wskaźnik bogactwa leksykalnego (I_{bog})
1.	Janiszewski M., <i>Dom nad rzeką Loes</i>	3396	10 000	6,8
2.	Górecki W., <i>Abchazja</i>	3273	10 000	6,5
3.	Bator J., <i>Japoński wachlarz</i>	3201	10 000	6,4
4.	Kęskrawiec M., <i>Czwarty pożar Teheranu</i>	3124	10 000	6,2
5.	Pawlikowska B., <i>Blondynka na Kubie. Na tropach prawdy i Ernesta Che Guevary</i>	3047	10 000	6,1
6.	Stasiuk A., <i>Jadąc do Babadag</i>	3014	10 000	6,0
7.	Cegielski M., <i>Oko świata. Od Konstantynopola do Stambułu</i>	3002	10 000	6,0
8.	Milewski J., <i>Dym się rozwiera</i>	3001	10 000	6,0
9.	Szczygieł M., <i>Gottland</i>	2959	10 000	5,9
10.	Kopczyński K., Sajewicz A., <i>Dybuk. Opowieść o nieważności świata</i>	2957	10 000	5,9
11.	Smoleński P., <i>Izrael już nie frunie</i>	2945	10 000	5,9
12.	Piskała K., <i>Dryland</i>	2945	10 000	5,9
13.	Pajęczkowska A., <i>Wędrowny zakład fotograficzny</i>	2876	10 000	5,8
14.	Wanat E., <i>Deutsche nasz. Reportaże berlińskie</i>	2875	10 000	5,8
15.	Szabłowski W., <i>Zabójca z miasta moreli. Reportaże z Turcji</i>	2793	10 000	5,6
16.	Wlekły M., <i>Raban. O kościele nie z tej ziemi</i>	2774	10 000	5,5
17.	Mikołajewski J., <i>Wielki przyptyw</i>	2631	10 000	5,3
18.	Rosiak D., <i>Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan</i>	2577	10 000	5,2
Średnia dla reportażu książkowego		2966	10 000	5,9

Źródło: opracowanie własne.

Spośród badanych tekstów z wartością $I_{\text{bog}} \geq 6,4$ trzy, tj. *Dom nad rzeką Loes, Japoński wachlarz, Abchazja*, wyróżniają się wysokim zintelektualizowaniem, wątkami przyczynowo-skutkowymi, prezentacją nie tylko reporterskich opisów podróźniczych, lecz także faktów, dat, ciekawostek historycznych charakterystycznych dla danej kultury, np.:

[1] W latach indonezyjskiej okupacji przeciwko niesławnym **samolotom Bronco OV-10**, skonstruowanym z myślą o zwalczaniu partyzanckich ugrupowań i podarowanym wojskowej dyktaturze przez **Amerykanów**, partyzanci z **FALINTIL-u** (Forças Armadas de Libertação Nacional Timor-Leste – Narodowych Wyzwoleńczych Sił Zbrojnych **Timoru Wschodniego**) budowali doły z powbijanymi na sztorc bambusami i pułapki z karabinowych kul (Janiszewski 2014: 22).

[2] Mimo iż w ciągu **XX wieku Tokio** dwukrotnie zostało całkowicie zniszczone, to wielkie trzęsienie ziemi z **1923 roku** nie było tak tragiczne w skutkach jak amerykańskie bomby w **1945 roku**. O ile bowiem **Japończycy** przywykli do trzęsień ziemi, o tyle nie mieli wcześniej okazji, by się przyzwyczaić do tak wielkich klęsk wojennych i obecności okupanta w swojej ojczyźnie (Bator 2004: 46).

[3] Abchazja próbuje przeciągnąć na swoją stronę siedemdziesiąt tysięcy galskich **Megrelów** i chce ich skłócić z **Tbilisi**. W **Suchumi** wychodzą prace o tym, że Megrelowie to oddzielny naród, prześladowany wiekami przez **Gruzinów** – przykładem książka **Aleksieja Arguna** *Nie nado unizat' swanow i mingriełow...* (Nie należy poniżać **Swanów** i Megrelów...). W **Gali** wznowiła działalność gazeta „Gał”, wychodząca po abchasku, rosyjsku i megrelsku. **Kiszmaria** realizuje ten mały abchaski prometeizm w terenie. Do tego potrzebne są pieniądze, których **Suchumi** nie przysyła. Gubernator musi sobie jakoś radzić (Górecki 2013: 40).

W powyższych fragmentach należy zauważyć mnogość dat (*XX wiek, 1923 rok, 1945 rok*), specjalistycznych terminów (*Bronco OV-10, FALINTIL*), nazw geograficznych (*Timor Wschodni, Tokio, Tbilisi, Gali, Suchumi*), nazw osobowych (*Aleksiej Argun, Kiszmaria*), nazw narodów (*Amerykanie, Japończycy, Megrelowie, Swanowie, Gruzińi*), zdań wielokrotnie złożonych przytaczających ważne polityczno-społeczne i ekonomiczne wydarzenia. Oznacza to, że reportaże z wysokim współczynnikiem ilościowego bogactwa słownictwa eksponują elementy charakterystyczne dla polszczyzny pisanej, rozwijają relacje statyczne (międzyprzedmiotowe) i eksponują tym samym dokumentarność gatunku.

Natomiast trzy teksty z najniższym badanym parametrem $I_{\text{bog}} \leq 5,5$, tj. *Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan, Wielki przyptyw, Raban. O kościele nie z tej ziemi*, nadal plasują się wśród tekstów pisanych, jednak zawierają znacznie więcej elementów charakterystycznych dla polszczyzny mówionej, np.:

[4] 5 stycznia, Antiochia

Deszcz. Ciągle pada... **A co ma robić deszcz w styczniu na granicy południowej**

Anatolii i Lewantu? Rzeka Orontes zmienia się w rwący potok, drogi spływają błotem, chodniki w miastach pękają pod naporem wody. [...]

Po wyjeździe z terminalu **mijamy** szkielet niedokończonego budynku, być może drugiego terminalu, który w połowie budowy okazał się zbędny. **Opuszczamy** lotnisko, a tam jak na Mazowszu: **płasko, ziemia zaorana, wszystko w ciapki. Pole kapusty, meczet, stacja benzynowa, sklep meblowy, osiedle bloków, składowisko złomu, restauracja, salon Volkswagena, szkoła policyjna...** (Rosiak 2015: 30).

[5] **Wracam. Idę** tam, gdzie za znakiem stoi „Lady Angela”. **Maleńka, drewniana łódka, błękitna** jak sprana sukienka dziewczynki, która dzisiaj jest babcią. Idę w stronę miasta. Po lewej wszędzie morze. **Dalekie, błękitne, jak z widokówek.** I stragany biednego targowiska za cementowym murkiem z napisem Buongiorno vita, dzień dobry, życie. Po prawej stronie mijam **półpiętrowe domki z płaskimi dachami i supermarket SISA.** Aż **dostrzegam** budynek, który musi być szpitalem, bo stoją przed nim dwa ambulanse.

Wysoki, dość masywny, w nowoczesnym stylu, którego nie umiem określić inaczej niż funkcjonalny (Mikołajewski 2015: 15).

[6] Argentyna

– Tu Jorge – zadzwonił kiedyś do Franciszka, księdza z Buenos Aires.

– **Nie pieprz** – odpowiedział Franciszek. Myślał, że ktoś sobie żartuje, i zaraz się rozłączył.

Zadzwonił raz jeszcze:

– Franciszku, to naprawdę ja, kardynał.

Kardynał dobrze **wiedział**, jaki z Franciszka gagatek. Kiedy policja niesłusznie aresztuje ludzi z dzielnicy, on **idzie się o nich wyklócać; urzęda** demonstracje, **pisze** listy do prezydenta, **nie boi się** przyjąć lesbijkę na wolontariuszkę, koleguje się z osobami transpłciowymi. Sam siebie nazywa księdzem awanturnikiem (Wleki 2019: 16).

Powyższe wypowiedzi cechują się występowaniem pytań retorycznych (*A co ma robić deszcz w styczniu na granicy południowej Anatolii i Lewantu?*), wyliczeń (*Pole kapusty, meczet, stacja benzynowa, sklep meblowy, osiedle bloków, składowisko złomu, restauracja, salon Volkswagena, szkoła policyjna...*), wulgaryzmów (*nie pieprz*) oraz znacznym dynamizmem, żywością wypowiedzi, co zostało zrealizowane poprzez nagromadzenie czasowników w formie osobowej (*mijamy, opuszczamy, wracam, idę, dostrzegam, wiedział, idzie się wyklócać, pisze, nie boi się*).

Warto przyjrzeć się innym typom tekstów, aby uzyskać szerszy pogląd na różne gatunki pisane i mówione. Średnia badanego współczynnika dla reportażu książkowego wynosi 5,9, natomiast dane dla tekstów naukowych zostały obliczone na podstawie badań Marii Rachwałowej (1986: 24), dla tekstów mówionych zaś – na bazie analiz Marii Zarębiny (1985: 130).

Tabela 2. Wartość wskaźnika ilościowego bogactwa słownictwa reportażu książkowego na tle porównawczym 14 różnych typów tekstów.

Lp.	Typ tekstu	Wskaźnik ilościowego bogactwa słownictwa (I_{bog})
1.	prace literaturoznawcze	6,1
2.	reportaż książkowy	5,9
3.	prace historyczne	5,7
4.	prace językoznawcze	5,4
5.	prace pedagogiczne	4,8
6.	audycje radiowe i telewizyjne	4,5
7.	język dyskusji naukowych	4,5
8.	prace naukowo-techniczne	4,5
9.	dialogi studenckie	4,4
10.	język w środowisku szkolnym (dialogi nauczyciel–uczeń)	4,2
11.	rozmowy potoczne	4,1
12.	język narad	3,7
13.	rozmowy telefoniczne	3,3
14.	język młodzieży licealnej	3,2
15.	język biurowy	3,1
Średnia		4,5

Źródło: opracowanie własne.

Przytoczone typy tekstów można zatem podzielić na trzy podstawowe grupy.

Pierwszą grupę stanowią teksty pisane, w których wartość wskaźnika ilościowego bogactwa słownictwa wynosi minimum 5,0 ($I_{bog} \geq 5,0$). Należą tutaj prace literaturoznawcze, reportaż książkowy, prace historyczne, prace językoznawcze. Reportaż książkowy ze względu na badany parametr zbliża się do humanistycznych dyscyplin naukowych, rozwija relacje statyczne (międzyprzedmiotowe), co wpływa na intelektualizację tekstu.

Do drugiej grupy zalicza się teksty z wartością I_{bog} większą lub równą 4,5 oraz mniejszą niż 5,0 ($4,5 \leq I_{bog} < 5,0$). Znajdują się tu zarówno naukowe teksty pisane z zakresu pedagogiki, prace naukowo-techniczne, jak i staranne wypowiedzi mówione, czyli audycje radiowe i telewizyjne, a także język dyskusji naukowych.

Trzecia grupa z wartością I_{bog} poniżej 4,5 ($I_{bog} < 4,5$) jest najbardziej charakterystyczna dla spontanicznych wypowiedzi mówionych. Występują tutaj dialogi studenckie, język w środowisku szkolnym (dialogi nauczyciel–uczeń), rozmowy potoczne, język narad, rozmowy telefoniczne, język młodzieży licealnej, język biurowy.

Wnioski

Podsumowując powyższe analizy, należy podkreślić kilka ważnych zależności.

Po pierwsze, pomiędzy analizowanymi reportażami książkowymi występuje znaczna różnica we wskaźniku bogactwa ilościowego słownictwa, która wynosi aż 1,6. Należy przypuszczać, że wynika ona ze stylu określonych reportaży, rozpiętych między polszczyzną pisaną a mówioną oraz między relacjami statycznymi i dynamicznymi.

Po drugie, styl reportażu książkowego mieści się w grupie najbardziej zintelektualizowanej ($I_{\text{bog}} \geq 5,0$), obok stylu tekstów literaturoznawczych, historycznych, językoznawczych. Potwierdza to tezę o stylistycznej bliskości reportażu książkowego i humanistycznych tekstów naukowych (zob. Ostrowska 2023a: 215–218).

I_{bog} sporo mówi o cechach stylistycznych analizowanych tekstów, związanych z użyciem leksyki. Jednak pozwala porównywać je wyłącznie ze względu na mniejsze lub większe podobieństwo ilościowe dotyczące bogactwa słownictwa. Tym samym „nie ma nic wspólnego z ocenami estetycznymi tekstów” (Sambor 1972: 219). Analizowane próby statystyczne muszą być jednakowej lub bardzo podobnej długości, by porównywanie wartości badanego parametru statystycznego miało sens. Jak się wydaje, warunek ten „rzadko bywa w praktyce badawczej spełniany” (Sambor 1972: 219).

Bibliografia podmiotowa

- Bator J., 2004, *Japoński wachlarz*, Warszawa.
- Cegielski M., 2009, *Oko świata. Od Konstantynopola do Stambułu*, Warszawa.
- Górecki W., 2013, *Abchazja*, Wołowiec.
- Janiszewski M., 2014, *Dom nad rzeką Loes*, Wołowiec.
- Kęskrawiec M., 2010, *Czwarty pożar Teheranu*, Warszawa.
- Kopczyński K., Sajewicz A., 2017, *Dybuk. Opowieść o nieważności świata*, Warszawa.
- Mikołajewski J., 2015, *Wielki przyptyw*, Warszawa.
- Milewski J., 2008, *Dym się rozwiera*, Poznań.
- Pajęczkowska A., 2019, *Wędrowny zakład fotograficzny*, Wołowiec.
- Pawlikowska B., 2006, *Blondynka na Kubie. Na tropach prawdy i Ernesta Che Guevary*, Warszawa.
- Piskała K., 2014, *Dryland*, Warszawa.
- Rosiak D., 2015, *Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan*, Wołowiec.
- Smoleński P., 2006, *Izrael już nie frunie*, Wołowiec.
- Stasiuk A., 2004, *Jadąc do Babadag*, Wołowiec.
- Szabłowski W., 2010, *Zabójca z miasta moreli. Reportaże z Turcji*, Wołowiec.
- Szczygieł M., 2006, *Gottland*, Wołowiec.
- Wanat E., 2018, *Deutsche nasz. Reportaże berlińskie*, Warszawa.
- Wlekły M., 2019, *Raban. O kościele nie z tej ziemi*, Warszawa.

Bibliografia przedmiotowa

- Gąsior K., 1997, *Od słowa do tekstu. O języku i stylach Janusza Korczaka*, Kraków.
- Guiraud P., 1954, *Les caractères statistiques du vocabulaire*, Paris.
- Guiraud P., 1966, *Zagadnienia i metody statystyki językoznawczej*, tłum. M. Kniaginina, Warszawa.
- Hammerl R., Sambor J., 1990, *Statystyka dla językoznawców*, Warszawa.
- Kuraszkiewicz W., 1958, *Statystyczne badanie słownictwa polskich tekstów XVI w.*, [w:] *Z polskich studiów slawistycznych*, seria 1, t. 1, Warszawa, s. 240–257.
- Laskowska E., 1985, *Z badań nad słownictwem uczniów i nauczycieli*, „Zeszyty Naukowe Wyższej Szkoły Pedagogicznej w Bydgoszczy. Studia Filologiczne. Filologia Polska” 24 (9), s. 151–163.
- Łojek M., 2001, *Słownictwo homilii w świetle badań statystycznych*, „Poradnik Językowy” 6, s. 32–48.
- Maryl M., Piasecki M., Walkowiak T., *Literacki Eksplorator Maszynowy*, CLARIN-PL, <https://ws.clarin-pl.eu/> (dostęp: 11.01.2024).
- Mayenowa M., 1965, *Możliwości i niebezpieczeństwa metod matematycznych w poetyce* [w:] *Poetyka i matematyka*, red. M. Mayenowa, Warszawa, s. 5–20.
- Mistrik J., 1967, *Matematiko-statisticzeskije metody v stilistike*, „Voprosy Jazykoznanija” 3, s. 41–50.
- Ostrowska K., 2023a, *Reportaż książkowy w świetle językoznawstwa statystycznego. Składnia, styl, gatunek*, Kielce.
- Ostrowska K., 2023b, *Słownictwo wybranych polskich reportaży książkowych (2004–2018)*, t. 1: *Listy frekwencyjne*, Kielce.
- Ostrowska K., Ruszkowski M., 2024, *Wskaźnik nominalności jako element stylistycznej charakterystyki współczesnego polskiego reportażu książkowego*, „Respectus Philologicus” s. 38–48.
- Pawłowski A., 2001, *Metody kwantytatywne w sekwencyjnej analizie tekstu*, Warszawa.
- Pisarek W., 1969, *Jak mierzyć zrozumiałość tekstu*, „Zeszyty Prasoznawcze” X, z. 4, s. 35–48.
- Pisarek W., 1972, *Frekwencja wyrazów w prasie. Wiadomości – komentarze – reportaże*, Kraków.
- Rachwałowa M., 1986, *Słownictwo tekstów naukowych*, Wrocław–Warszawa–Kraków–Gdańsk–Łódź.
- Ruszkowski M., 2004a, *Wskaźnik epitetyzacji w badaniach stylistycznych*, „Respectus Philologicus” 5 (10), s. 48–53.
- Ruszkowski M., 2004b, *Wskaźnik nominalności jako element stylistycznej charakterystyki tekstu*, „Stylistyka” 13, s. 341–348.
- Ruszkowski M., 2022, *O wskaźniku wariantywności fleksyjnej*, [w:] *Polszczyzna w różnych jej aspektach*, red. A. Dyszak. *Język polski dawniej i dziś*, red. A. Paluszak-Bronka, Bydgoszcz, s. 229–236.
- Sambor J., 1972, *Słowa i liczby. Zagadnienia językoznawstwa statystycznego*, Warszawa.
- Sigurd B., 1975, *Struktura języka. Zagadnienia i metody językoznawstwa współczesnego*, tłum. Z. Wawrzyński, Warszawa.
- Smółkowa T., 1974, *Słownictwo i fleksja „Lalki” Bolesława Prusa. Badania statystyczne*, Wrocław–Warszawa–Kraków–Gdańsk.

- Vaškevič-Buš J., 2017, *Kompetencje leksykalne uczniów w młodszym wieku szkolnym w Polsce i na Litwie (studium statystyczno-kognitywne)*, Kraków.
- Wierzbicka A., 1987, *O języku dla wszystkich*, wyd. 2, Warszawa.
- Zarębina M., 1970, *Rola wyrazów w słowniku i tekście*, „Język Polski” L, s. 33–46.
- Zarębina M., 1971, *Najczęstsze wyrazy polszczyzny mówionej*, „Język Polski” LI, s. 336–347.
- Zarębina M., 1985, *Próba statystycznej analizy słownictwa polszczyzny mówionej (synteza danych liczbowych)*, Wrocław–Warszawa–Kraków–Gdańsk–Łódź.
- Zembaty-Michalakowa M., 1982, *Poezja Juliana Przybosa w świetle badań statystyczno-językowych na tle porównawczym*, Wrocław.
- Zgółkowska H., 1987, *Ilościowa charakterystyka słownictwa współczesnej polszczyzny. Wybrane zagadnienia*, Poznań.
- Zipf G., 1935, *The Psychobiology of Language*, Boston.

An indicator of the quantitative richness of vocabulary as an element of the lexical characteristics of contemporary Polish book reportage

Abstract

One of the most important parameters for characterizing the vocabulary of a text is the *index of quantitative vocabulary richness* (I_{bog}). The most frequently used indicator is developed by Jozef Mistrik: $I_{bog} = \frac{20V}{N}$ (V – the number of different lexemes in the text, i.e. the text dictionary; N – text length, i.e. the number of all word forms). Using J. Mistrik's formula (N – number of word forms, i.e. 10,000 from the beginning of each reportage, V – number of lexemes in each of these samples), I_{bog} was calculated for individual texts. The value of the indicator ranges between 5.2 (D. Rosiak, *Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan*) and 6.8 (M. Janiszewski, *Dom nad rzeką Loes*). There is, therefore, a significant difference in the quantitative vocabulary richness index among the analyzed book reports, which amounts to as much as 1.6. It should be assumed that it results from the style of specific reportages, spanning between written and spoken Polish and between static and dynamic relationships. The average for the reportage $I_{bog} \approx 5.9$. Therefore, due to the examined parameter, it is close to the humanistic scientific disciplines (literary, historical, linguistic works in which ranges between 5.4 and 6.1), which indicates a high lexical richness of book reportage.