

Opis leksyki współczesnych polskich reportaży książkowych na tle innych typów tekstów z wykorzystaniem wskaźnika powtarzalności słownictwa

The Comparison of Lexis in Contemporary Polish Book Reportages
and Other Types of Texts with the Use of the Vocabulary Repetition Index

Abstract: One of the parameters characterizing the lexis of a text is the vocabulary repetition index comparable only when we examine texts or representative samples of texts of equal length. To compare texts of different (or similar) stylistic provenance, the most common way is to analyze 10,000 word forms and examine the index value for this sample. The value of the index for the Polish book reportages analysed in the article ranges from 2.94 to 3.88. The average value for the reportages under scrutiny is 3.39. In terms of the value of the analyzed parameter, it is close to humanistic scientific projects, especially those in the fields of literary studies and history. It seems justified by the thematic scope of reportages, which often refers to historical facts, and the method of their literary presentation. The reason for this is also because in reportages a large group of words is independent from the author, for example, place names, surnames, political science terms, etc. The research shows that the level of vocabulary repetition in book reportages is very low. Consequently, book reportages are characterized by diverse, rich and original lexis.

Key words: quantitative analysis, frequency of word forms, frequency of lexemes, repetition of vocabulary, book reportage

Abstrakt: Jednym z parametrów charakteryzujących leksykę tekstu jest wskaźnik powtarzalności słownictwa, porównywalny tylko wtedy, gdy badamy teksty lub reprezentatywne próbki tekstów o jednakowej długości. Aby porównywać teksty o różnej (lub podobnej) stylistycznej proveniencji, najczęściej analizie poddaje się 10 tys. słowoform i bada wartość wskaźnika dla tej próby. Wartość wskaźnika dla badanych w artykule polskich reportaży książkowych waha się między 2,94 a 3,88. Średnia wartość wskaźnika dla badanych reportaży wynosi ok. 3,39 i jest zbliżona do wartości analizowanego parametru dla humanistycznych prac naukowych, głównie z zakresu literaturoznawstwa i historii. Wydaje się to uzasadnione tematyką reportaży, odwołujących się często do faktów historycznych, i sposobem literackiej prezentacji tekstów, a także tym, że w reportażach dużą grupę stanowią wyrazy niezależne od autora (nazwy miejscowe, nazwiska, terminy politologiczne itp.).

Badania dowodzą, że reportaż książkowy cechuje się małą powtarzalnością słownictwa, a co za tym idzie – leksyką zróżnicowaną, bogatą i oryginalną.

Słowa kluczowe: analiza kwantytatywna, frekwencja słowoform, frekwencja leksemów, powtarzalność słownictwa, reportaż książkowy

Wprowadzenie

Zdaniem Witolda Mańczaka (1996) „problemy roztrząsane przez językoznawców rozpadają się przede wszystkim na dwie kategorie: 1) te, które można rozstrzygać za pomocą statystyki, oraz 2) te, których za pomocą statystyki rozstrzygnąć się nie da” (s. 8)¹. Różnorodne wskaźniki ilościowe – nominalności², epitetyzacji³, bogactwa językowego, powtarzalności słownictwa – umożliwiają przeanalizowanie stylistycznych wyznaczników określonych tekstów; na podstawie tych wyznaczników możliwe jest formułowanie wniosków na temat cech językowych gatunków literackich, w tym gatunku reportażu książkowego.

Reportaż książkowy rozumiany jest w niniejszym artykule jako gatunek wypowiedzi, stanowiący – zgodnie z genologią lingwistyczną (zob. Wojtak, 2019, s. 114–121) – wzorzec gatunkowy, a jednocześnie spełniający cztery podstawowe kryteria, takie jak:

- struktura – oparty na faktach model kompozycyjny tzw. noweli dokumentarnej;
- pragmatyka – relacje nadawczo-odbiorcze pomiędzy reporterem a czytelnikiem;
- tematyka – treści dotyczące aktualnych wojennych, politycznych, religijnych, kulturowych problemów na świecie;
- stylistyka – „językowe ukształtowanie łączące w sobie elementy stylów funkcjonalnych (artystycznego, publicystycznego, naukowego) oraz zawierające charakterystyczne stylistyczno-syntaktyczne wykładniki, tworzące tym samym indywidualny styl reportażowy” (Ostrowska, 2023, s. 19–20).

Warto podkreślić, że konsekwentnie używana jest tutaj nazwa „reportaż książkowy” – zamiast nazwy „reportaż literacki”, odnoszącej się do gatunku rozpowszechnionego i często utożsamianego z formą monograficzną – ze względu na to, że przymiotnik „literacki” prowadzi na myśl styl literacki. Reportaż w formie książkowej posiada natomiast wykładniki różnorodnych stylów funkcjonalnych, w tym publicystycznego, naukowego i literackiego. Dodatkowo literackość niejednokrotnie utożsamiana jest z fikcją, która w reportażu nie powinna mieć miejsca.

Wskaźnik powtarzalności słownictwa (W_{pow}) nazywany jest także wskaźnikiem średniej częstości wyrazu (Rachwałowa, 1986, s. 25) lub wskaźnikiem iteracji (Sambor, 1972, s. 221). Jako jeden z wielu wskaźników umożliwia wnioskowanie o cechach leksykalnych

¹ Maria Mayenowa (1965) zwracała uwagę na to, że „intuicją statystyczną posługują się nawet ci, którzy na teorii są jej gorącymi przeciwnikami” (s. 14). Adam Pawłowski (2001) podkreśla zaś, że lingwistyka kwantytatywna zakłada „mierzalność i/lub kwantyfikowalność pewnych cech językowych” (s. 6).

² Za pośrednictwem wskaźnika nominalności można wnioskować o rozwijaniu relacji statycznych bądź dynamicznych, parametr ten świadczy również o żywości określonych tekstów lub gatunków.

³ Wskaźnik epitetyzacji może świadczyć o stopniu oficjalności, półoficjalności, nieoficjalności danego tekstu i stanowić narzędzie do orzekania o potoczności danego stylu czy języka.

analizowanych tekstów, natomiast pozwala je porównywać tylko ze względu na mniejszą lub większą powtarzalność słownictwa. Tym samym „nie ma nic wspólnego z ocenami estetycznymi tekstów” (Sambor, 1972, s. 219), których cechą stylistycznie relewantną może być celowa asceza leksykalna. Im wartość tego wskaźnika jest mniejsza, tym mniejsza jest powtarzalność słownictwa, czyli zwiększa się bogactwo i zróżnicowanie leksyki. Jeżeli wartość wskaźnika jest większa, tekst charakteryzuje się uboższym słownictwem, a wyrazy są częściej powtarzane. Omawiany wskaźnik oblicza się ze wzoru:

$$W_{pow} = \frac{N}{V},$$

gdzie:

W_{pow} to wartość wskaźnika powtarzalności słownictwa;

N – liczba wszystkich słowoform;

V – liczba różnych leksemów w tekście.

Należy pamiętać o tym, że wyniki badania wskaźnika są miarodajne wyłącznie wtedy, gdy porównujemy 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.

Jak pisali Maria Kniaginina i Walery Pisarek (1966, s. 15–16), słownictwo można podzielić na dwie grupy – wyrazy zależne i niezależne od autora tekstu. O użyciu pierwszych decyduje autor, który ma do wyboru liczne synonimy; drugich co do zasady musi użyć, ponieważ wyrazy te są zdeterminowane przez rzeczywistość pozajęzykową. Do drugiej grupy należą: nazwy geograficzne, określenia dat, imiona, nazwiska, nazwy instytucji, organizacji, nazwy zawodów, tytułów, wydarzeń historycznych, wyrobów przemysłowych, terminy politologiczne, których w reportażach jest dużo. Jak wykazują badania, już same nazwy własne to ok. 22% słownictwa analizowanych w tym badaniu reportaży – obejmują 10 330 leksemów ze wszystkich 47 020. Można zatem przypuszczać, że reportaż książkowy ma bardzo niski wskaźnik powtarzalności słownictwa ze względu na szeroką gamę różnorodnych wyrazów zarówno tych zależnych od autora, które umożliwiają konstruowanie myśli z użyciem synonimów, jak i niezależnych od autora, w postaci np. toponimów, wzmagających dokumentarność tego gatunku.

Założenia artykułu

Celem artykułu jest uzyskanie odpowiedzi na dwa pytania.

1. Czy w polskich reportażach książkowych z badanego korpusu średni wskaźnik powtarzalności słownictwa jest wysoki?
2. Jaki jest średni wskaźnik powtarzalności słownictwa polskiego reportażu książkowego na tle wskaźników odnoszących się do innych typów tekstów?

Aby odpowiedzieć na te pytania, przeprowadzono badania, w których posłużono się kilkoma metodami: korpusową, analizy ilościowej oraz analizy porównawczej.

W pierwszej kolejności zastosowano metodę korpusową – indukcyjne podejście oparte na korpusie (ang. *corpus-driven approach*), uważane za autonomiczny paradygmat językoznawczy, w ramach którego „wszelkie hipotezy badawcze formułuje się właśnie w oparciu o częstość występowania i dystrybucję wyrazów lub ich sekwencji w tekstach (np. n-gramów, ram frazowych)” (Grabowski i Hebal-Jezierska, 2016, s. 68–69). Do korpusu tekstów zaliczono utwory spełniające jednocześnie wymogi: genologiczny (teksty należące do gatunku wypowiedzi, jakim jest reportaż, stanowiące wzorzec gatunkowy, w którym brane są pod uwagę aspekty takie jak: struktura, pragmatyka, tematyka, stylistyka), języka (napisane w języku polskim), formy (wydane w formie książkowej), czasu (reportaże współczesne).

Wymienione warunki zostały spełnione przez współczesne, XXI-wieczne, polskie reportaże książkowe wyróżnione Nagrodą im. Beaty Pawlak. Do korpusu badanych tekstów nie włączono reportaży publicystycznych wydanych w pierwotnej formie w czasopiśmie; zaklasyfikowano zaś 18 wybranych tekstów z lat 2004–2019, tj.:

1. *Jadąc do Babadag* Andrzeja Stasiuka (2004).
2. *Japoński wachlarz* Joanny Bator (2004).
3. *Blondynka na Kubie. Na tropach prawdy i Ernesta Che Guevary* Beaty Pawlikowskiej (2006).
4. *Gottland* Mariusza Szczygła (2006).
5. *Izrael już nie frunie* Pawła Smoleńskiego (2006).
6. *Dym się rozwiewa* Jacka Milewskiego (2008).
7. *Oko świata. Od Konstantynopola do Stambułu* Mała Cegielskiego (2009).
8. *Czwarty pożar Teheranu* Marka Kęskrawca (2010).
9. *Zabójca z miasta moreli. Reportaże z Turcji* Witolda Szablowskiego (2010).
10. *Abchazja* Wojciecha Góreckiego (2013).
11. *Dom nad rzeką Loes* Mateusza Janiszewskiego (2014).
12. *Dryland* Konrada Piskąły (2014).
13. *Wielki przypływ* Jarosława Mikołajewskiego (2015).
14. *Ziarno i krew* Dariusza Rosiaka (2015).
15. *Dybuk. Opowieść o nieważności świata* Krzysztofa Kopczyńskiego i Anny Sajewicz (2017).
16. *Deutsche nasz. Reportaże berlińskie* Ewy Wanat (2018).
17. *Raban. O kościele nie z tej ziemi* Mirosława Wlekłego (2019).
18. *Wędrowny zakład fotograficzny* Agnieszki Pajączkowskiej (2019).

Aby spełnić warunek porównywalności wskaźnika oraz jednakowej wielkości próby, z początku każdego z tekstów wyekscerpowano próbki równo po 10 tys. słowoform (N); dokonano tego metodami wykorzystywanymi w badaniach Marii Rachwałowej oraz Marii Zarębiny. Za pomocą Literackiego Eksploratora Maszynowego (LEM) z infrastruktury CLARIN-PL reportaże poddano analizie ilościowej, tj. zlematyzowano korpus, wyznaczono statystykę leksemów i słowoform oraz obliczono wskaźnik powtarzalności dla poszczególnych reportaży, korzystając ze wzoru Rachwałowej – gdzie N to liczba słowoform, tu: po 10 tys. słowoform z początku każdego reportażu, a V to liczba leksemów w każdej z tych prób.

Ostatnim etapem badań było dokonanie analizy porównawczej wartości wskaźnika W_{pow} polskiego reportażu książkowego z wartościami wskaźników dotyczącymi słownictwa innych gatunków tekstów, w tym polszczyzny mówionej (Zarębina, 1985, s. 65) oraz tekstów

naukowych (Rachwałowa, 1986, s. 26), analizowano przy tym język prac literaturoznawczych, historycznych, językoznawczych, pedagogicznych, naukowo-technicznych, audycji radiowych, telewizyjnych, dyskusji naukowych, narad, rozmów telefonicznych, młodzieży licealnej, biurowy, dialogów studenckich, dialogów nauczyciel – uczeń, rozmów potocznych. Wartości wskaźnika powtarzalności słownictwa są wiarygodne, ponieważ liczba słowoform w porównywanych pracach jest zbliżona do 10 tys. (waha się pomiędzy 9 972 a 10 030 słowoform).

Wskaźnik powtarzalności słownictwa w badanych reportażach

Pomiędzy badanymi tekstami zdaniem autorów tego artykułu występują znaczne różnice w zakresie powtarzalności słownictwa. Rozpiętość między wskaźnikami dotyczącymi polszczyzny mówionej i pisanej wynosi 1,1, w reportażach natomiast waha się między wartością 2,94 (Janiszewski, 2014) a 3,88 (Rosiak, 2015), czyli wynosi 0,94. Powtarzalność słownictwa w reportażu Janiszewskiego jest najmniejsza (czyli zróżnicowanie leksyki największe), a w tekście Rosiaka – powtarzalność słownictwa jest największa (czyli bogactwo leksyki najmniejsze). Średnia dla wszystkich badanych reportaży wynosi $W_{pow} \approx 3,39$. Dla porównania należy dodać, że wartość W_{pow} w polszczyźnie pisanej wynosi 3,04 (Zarębina, 1970, s. 33–46), a w polszczyźnie mówionej – 4,14 (Zarębina, 1971, s. 336–347), jest więc znacząco wyższa. Reportaż zbliża się pod tym względem do polszczyzny pisanej, co wydaje się oczywiste, biorąc pod uwagę fakt, że reportaż to pisana odmiana polszczyzny. Język mówiony charakteryzuje się zdecydowanie większą powtarzalnością słownictwa, czyli uboższą i mniej oryginalną leksyką.

Reportaż o najwyższym wskaźniku powtarzalności i jednocześnie najmniejszym parametrze bogactwa językowego to *Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan* Rosiaka ($W_{pow} = 3,88$). Na 10 tys. słowoform w obrębie nazw własnych występuje tutaj 230 lematów z łączną frekwencją aż 561 wystąpień. W tekście reportażowym zaś słowami kluczowymi, które zaliczają się do słownictwa niezależnego od autora, są m.in.: *Izrael, Liban, Syria, Turcja, Egipt, Palestyńczyk, Jerozolima*, np.: „Remon wyemigrował w 1976 roku, podczas wojny Syrii z Izraelem” (Rosiak, 2015, s. 50); „Często słyszałem też plotkę, że Mursi w zamian za urząd prezydenta (wybory były sfałszowane za zgodą Amerykanów – tak głosi plotka) zawarł z prezydentem Obamą umowę, że odda północny Synaj Hamasowi i w ten sposób Izrael będzie mógł się pozbyć Palestyńczyków” (s. 62); „Został księdzem w 1964 roku, wyjechał na kilka lat do Jerozolimy, od roku 1970 przez dwadzieścia osiem lat wykładał na uniwersytecie w Kasliku, tam gdzie dziś nauczają Gabi” (s. 75); „Pochodzi z Midyat na południu Turcji, miał zaledwie sześć miesięcy, gdy rodzice uciekli do Niemiec tuż po wybuchu konfliktu grecko-tureckiego o Cypr” (s. 89); „Kair jest stolicą Egiptu od ponad tysiąca lat, ale ma w sobie tradycję poprzednich stolic i ośrodków egipskiej cywilizacji” (s. 102).

Tabela 1
Wartości wskaźnika powtarzalności słownictwa dla wybranych polskich reportaży książkowych – uporządkowanych od najwyższej wartości wskaźnika do najniższej (dla N = 10 000)

Autor i tytuł reportażu	Wielkość słownika / liczba leksemów (V)	Wskaźnik powtarzalności słownictwa (W_{pow})
D. Rosiak, <i>Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan</i>	2 577	3,88
J. Mikołajewski, <i>Wielki przyptyw</i>	2 631	3,80
M. Wlekły, <i>Raban. O kościele nie z tej ziemi</i>	2 774	3,60
W. Szablowski, <i>Zabójca z miasta moreli. Reportaże z Turcji</i>	2 793	3,58
E. Wanat, <i>Deutsche nasz. Reportaże berlińskie</i>	2 875	3,48
A. Pajączkowska, <i>Wędrowni zakład fotograficzny</i>	2 876	3,48
K. Piskala, <i>Dryland</i>	2 945	3,40
P. Smoleński, <i>Izrael już nie frunie</i>	2 945	3,40
K. Kopczyński, A. Sajewicz, <i>Dybuk. Opowieść o nieważności świata</i>	2 957	3,38
M. Szczygieł, <i>Gottland</i>	2 959	3,38
J. Milewski, <i>Dym się rozwiewa</i>	3 001	3,33
M. Cegielski, <i>Oko świata. Od Konstantynopola do Stambułu</i>	3 002	3,33
A. Stasiuk, <i>Jadąc do Babadag</i>	3 014	3,32
B. Pawlikowska, <i>Blondynka na Kubie. Na tropach prawdy i Ernesta Che Guevary</i>	3 047	3,28
M. Kęskrawiec, <i>Czwarty pożar Teheranu</i>	3 124	3,20
J. Bator, <i>Japoński wachlarz</i>	3 201	3,12
W. Górecki, <i>Abchazja</i>	3 273	3,06
M. Janiszewski, <i>Dom nad rzeką Loes</i>	3 396	2,94
Średnia	2 966	3,39

Adnotacja. N – liczba słowoform.

Wyrazy niezależne od autora występują przy okazji przytaczania ważnych wątków wojennych (dotyczących wojny Syrii z Izraelem, konfliktu grecko-tureckiego o Cypr), politycznych (jak wybory) czy historycznych (sytuacji, gdy Kair był stolicą Egiptu). Oprócz toponimów (tj. *Syria, Izrael, Synaj, Jerozolima, Kaslik, Midyat, Turcja, Niemcy, Cypr, Kair, Egipt*) w przytoczonych wypowiedzeniach występują nazwy mieszkańców państw (np. *Amerykanie, Palestyńczycy*), nazwiska (np. *Obama*), daty (tj. *1964 rok, 1970 rok, 1976 rok*).

Z kolei tekst reportażowy *Dom nad rzeką* Loes Janiszewskiego (2014) cechuje się najniższym wskaźnikiem powtarzalności słownictwa ($W_{pow} = 2,94$), co oznacza, że jest to najbardziej oryginalny pod względem leksykalnym tekst, czyli charakteryzujący się zróżnicowaną leksyką. W badanej próbie 10 tys. słowoform występuje 237 lematów w postaci nazw własnych z łączną frekwencją równą 358, czyli wystąpień jest tutaj o 203 mniej niż w tekście Rosiaka (2015). Wśród leksemów kluczowych należących do warstwy wyrazów niezależnych od autora można wyróżnić m.in. te: *Timor* ($f = 85$), *Indonezja* ($f = 17$), *ONZ* ($f = 16$), *Australia* ($f = 11$), *Indonezyjczyk* ($f = 9$). Warto podkreślić, że tego typu leksemów jest sporo, nie występują one jednak z tak dużą frekwencją jak te w tekście Rosiaka, np.: „Słuchają razem muzyki, jednocześnie takiej samej i innej niż ta, której słuchał na Timorze. Wychodzi na zewnątrz i już się wybudzając, zaczyna rozumieć, że jest w Indonezji” (Janiszewski, 2014, s. 93); „Czterysta kilometrów od północnych wybrzeży Australii i zgodnie z pierwotną umową, podpisaną jeszcze przez Indonezję i Australię, powinno być podzielone równo pomiędzy oba kraje” (s. 36); „Każdy zarzeka się na Najświętszą Panienkę, na pamięć przodków, na sekretarza generalnego ONZ, że to on mówi prawdę, że to jego wersja jest obiektywna” (s. 43); „»Tylko w cieśninie Wetar i jeszcze dwóch miejscach można przepłynąć z Pacyfiku na Ocean Indyjski, nie nadkładając drogi dookoła Australii i jednocześnie pozostając niezauważonym«– mówi mu wiele lat później emerytowany bosmanmat” (s. 77); „Dan przyjechał tu pierwszy raz w 1997 roku, jeszcze gdy byli tu Indonezyjczycy” (s. 25).

W przytoczonych wypowiedzeniach został wyeksponowany ważny dla reportażu wątek geograficzny (granice Timoru, cieśnina Wetar), przywołano daty (1997 rok, 1999 rok), toponimię (*Timor, Australia, Indonezja, Timor Wschodni, cieśnina Wetar, Pacyfik, Ocean Indyjski*). Pomimo że to słownictwo istotne, występuje w książce Janiszewskiego ze znacznie niższą frekwencją niż w innych tekstach.

Wskaźnik powtarzalności słownictwa na tle porównawczym

Wartość wskaźnika powtarzalności słownictwa dla reportażu książkowego porównano z wartościami wskaźników dla innych typów tekstów (tabela 2). Średnia wartość wskaźnika powtarzalności słownictwa wynosi 4,66. Rozpiętość pomiędzy najwyższą ($W_{pow} = 6,47$ – język biurowy) a najniższą wartością ($W_{pow} = 3,28$ – prace z zakresu literaturoznawstwa) wartością wskaźnika jest znaczna i wynosi 3,19.

Wskaźnik powtarzalności przyjmuje najwyższe wartości dla języka biurowego, młodzieżowego i rozmów telefonicznych. Może to wynikać z częstego stosowania w tych kontekstach utartych zwrotów, schematycznych wyrażen i ograniczonego zasobu słownictwa. Niższe wartości wskaźnika odnoszą się zaś do prac naukowych, literaturoznawczych i reportaży książkowych. Świadczy to o większej różnorodności leksykalnej tych tekstów, używaniu w nich bardziej złożonego i precyzyjnego języka oraz unikaniu powtórzeń.

Tabela 2
Wartość wskaźnika powtarzalności słownictwa dla tekstów różnych typów

Typ tekstu	Wskaźnik powtarzalności słownictwa (W_{pow})
Język biurowy*	6,47
Język młodzieży licealnej	6,28
Język rozmów telefonicznych	6,05
Język narad	5,39
Rozmowy potoczne	4,92
Język w środowisku szkolnym (dialogi nauczyciel – uczeń)	4,84
Dialogi studenckie	4,58
Prace naukowo-techniczne	4,47
Język dyskusji naukowych	4,44
Język audycji radiowych i telewizyjnych	4,42
Prace z zakresu pedagogiki	4,13
Prace z zakresu językoznawstwa	3,73
Prace z zakresu historii	3,49
Współczesny polski reportaż książkowy	3,39
Prace z zakresu literaturoznawstwa	3,28
Średnia	4,66

Adnotacja. Opracowanie własne na podstawie wyników badań językoznawców (m.in. Rachwałowa, 1986, s. 26; Zarębina, 1985, s. 65).

* Język biurowy rozumiany jest tutaj jako język, którym posługują się pracownicy administracyjni w komunikacji pomiędzy sobą. Charakteryzuje się specjalistyczną terminologią i najwyższym wskaźnikiem powtarzalności słownictwa, a tym samym najniższym zróżnicowaniem leksykalnym.

Analiza danych z tabeli 2 pozwala na podział tekstów na dwie podstawowe grupy. Pierwszą grupę stanowią te teksty, dla których wskaźnik powtarzalności jest wysoki i wynosi powyżej 4,5. To dialogi studenckie, język używany w środowisku szkolnym, rozmowy potoczne, język narad, rozmów telefonicznych, młodzieży licealnej oraz język biurowy. Oznacza to, że teksty mówione co do zasady obarczone są licznymi powtórzeniami, mniejszym zróżnicowaniem leksykalnym, anakolutami, a tym samym są nieoficjalne – niezależnie od sytuacji, która może wydawać się oficjalna, tak jak w przypadku pracy biurowej. Do drugiej grupy tekstów zalicza się natomiast te, dla których wartość badanego parametru jest mniejsza od 4,5 lub równa 4,5, czyli prace naukowo-techniczne, pedagogiczne, językoznawcze, literaturoznawcze, historyczne, język dyskusji naukowych, audycji radiowych i telewizyjnych, a także teksty reportażowe. W tej grupie znalazły się zatem przede wszystkim teksty pisane, ale również mówione charakteryzujące się oficjalnością,

wymagające rozległej wiedzy, uprzedniego skrupulatnego przygotowania i oparcia na faktach, datach, autorytetach.

Można zatem mówić o wyraźnej prawidłowości – im tekst jest swobodniejszy, mniej oficjalny, tym powtarzalność słownictwa jest większa. Dla tekstów pisanych wartość badanego parametru nie przekracza 4,5, podczas gdy dla tekstów mówionych wynosi nawet powyżej 6,0 (jak dla języka rozmów telefonicznych, młodzieży licealnej, polszczyzny używanej w biurze).

Reportaż książkowy zbliża się pod względem wartości analizowanego parametru ($W_{pow} = 3,39$) do humanistycznych prac naukowych, przede wszystkim z zakresu literaturoznawstwa ($W_{pow} = 3,28$) i historii ($W_{pow} = 3,49$). Tym samym kwalifikuje się do drugiej grupy, wskaźnik powtarzalności słownictwa w reportażach nie jest wysoki, teksty te charakteryzują się różnicowaniem leksykalnym oraz stylem oficjalnym. Jest to uzasadnione dwiema kwestiami, po pierwsze, tematyką reportaży, odwołującą się często do faktów historycznych, a także sposobem ich literackiej prezentacji, po drugie, wyznacznikami stylu naukowego, którymi są m.in. „rozbudowane, długie konstrukcje hipotaktyczne, hipotaktyczno-parataktyczne, złożone zestawione oraz wewnętrznie nawiązane” (Ostrowska, 2023, s. 205).

Reporter, przedstawiając określone sytuacje, dokonuje „transformacji dostrzeganych wydarzeń i faktów” (Wolny-Zmorzyński, 2004, s. 19), co stanowi w reportażach wyróżnik stylu artystycznego, obok występowania narratora, portretowania cech bohaterów, posługiwania się dialogiem i monologiem, plastyczności opisów. Obrazowość przedstawianych faktów to także ważna cecha literackości reportażu. Zdaniem Magdaleny Piechoty (2017, s. 348) jest ona „utrwalona poprzez jakości stylistyczne, ale ich estetyczny wymiar jest uzależniony od wybranej przez nadawcę roli i projektowanych celów reportażu (pragmatyka) oraz złożonych i faktycznych funkcji poznawczych przekazu”, co ilustruje przykład z reportażu Janiszewskiego (2014), w którym w barwny, obrazowy sposób przedstawiona jest sytuacja kontroli na granicy:

Dzisiaj kontrola na granicy, dziś tworzy się tkanę snów i raportów miesięcznych. Liczy głowy, mierzy przyrost miejscowych zdolności. Dzisiejsze *capacity* odjąć wczorajsze *capacity* – kontrolerzy mruczą święte słowo, które legitymizuje ich pobyt tutaj. Delta. Czyli zmiana. Wszystko po to, by stworzyć paliwo dla oenzetowskiej biurokratycznej maszyny, usprawiedliwić wypłacanie pensji, premii za pracę w strefie działań wojennych i zwrot kosztów podróży do tropikalnych rajów, gdzie można odreagować stres związany z ekspozycją na ludzkie nieszczęście. (s. 15)

Odwoływanie się do faktów historycznych jest zaś wykładnikiem stylu naukowego, cechującego się znacznym obiektywizmem, uporządkowaniem przyczynowo-skutkowym, spójnością, przywoływaniem określonych dat, np.:

Znalazłem się w światowej stolicy Asyryjczyków. Na dziewięćdziesiąt tysięcy mieszkańców Södertälje ponad połowa to obcokrajowcy, a trzydzieści tysięcy to chrześcijanie z Turcji, Iraku, Syrii. Osiedlają się w Szwecji od ponad czterdziestu

lat, a kolejne fale imigracji zbiegają się z datami bliskowschodnich konfliktów i ludzkich tragedii. Wojny państw arabskich z Izraelem, grecko-turecki konflikt o Cypr, rzezie w Libanie, rewolucja w Iranie i konflikt iracko-irański w latach osiemdziesiątych, pierwsza wojna w Zatoce Perskiej w latach dziewięćdziesiątych, potem inwazja na Irak i obalenie Saddama Husajna na początku XXI wieku, wreszcie od 2011 roku wojna domowa w Syrii – to najnowsze akty wielkiej tragedii, exodusu chrześcijan z miejsc, w których ich przodkowie żyli od tysięcy lat. Wcześniej też nie było im łatwo, ale trwali. Teraz stanęli wobec widma ostatecznej katastrofy. (Rosiak, 2015, s. 69)

Reportaż książkowy cechuje się niską powtarzalnością słownictwa (mniejsza wartość tego wskaźnika odnosi się tylko do prac literaturoznawczych), a co za tym idzie – zróżnicowanym, bogatym i oryginalnym słownictwem.

Wnioski

Przedstawiona w artykule analiza pozwala na odpowiedź na postawione na jego wstępie pytania.

Po pierwsze, średni wskaźnik powtarzalności słownictwa w polskich reportażach książkowych nie jest wysoki – wynosi 3,39, reportaże mieszczą się w grupie tekstów o niskiej wartości badanego parametru, co świadczy o oficjalności analizowanych tekstów reportażowych, ich dokumentarności, oparciu na faktach, historii, nauce, autorytetach. Warto zauważyć, że pomiędzy skrajnymi wartościami wskaźników powtarzalności słownictwa analizowanych reportaży książkowych występuje różnica 0,94 ($W_{pow} = 2,94$ dla tekstu Janiszewskiego pt. *Dom nad rzeką Loes*; $W_{pow} = 3,88$ dla reportażu Rosiaka pt. *Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan*). Należy przypuszczać, że wynika ona z dwóch kwestii: 1) tematyki tych reportaży – Rosiak pisze o Bliskim Wschodzie obejmującym kilkadziesiąt państw, Janiszewski zaś o jednym państwie; 2) leksykalnego ukształtowania reportażu, zakresu użycia wyrazów zależnych i niezależnych od autora – w reportażu Rosiaka frekwencja nazw własnych wynosi 561, a w tekście Janiszewskiego – 358.

Po drugie, średnia badanego parametru dla reportażu wynosi 3,39. Tak więc zbliża się on pod względem wartości wskaźnika powtarzalności słownictwa do humanistycznych prac naukowych, przede wszystkim z zakresu literaturoznawstwa ($W_{pow} = 3,28$) i historii ($W_{pow} = 3,49$). Wiąże się to z zakresem tematycznym reportażu, odwołującym się często do faktów historycznych (co jest wyróżnikiem stylu naukowego), oraz ze sposobem ich literackiej prezentacji (to z kolei wykładnik stylu artystycznego). Wynika to również z faktu, że w reportażach dużą grupę stanowią wyrazy niezależne od autora – to minimum 22% słownictwa analizowanych tekstów. Reportaż książkowy cechuje się więc bardzo małą powtarzalnością słownictwa i zróżnicowaną leksyką. Sytuacje opisywane w reportażach przedstawiane są w sposób barwny, obrazowy, plastyczny, reporter przywołuje konkretne daty, a także nazwy własne – należące do warstwy wyrazów niezależnych od autora.

Literatura

- Bator, J. (2004). *Japoński wachlarz*. Znak.
- Cegielski, M. (2009). *Oko świata. Od Konstantynopola do Stambułu*. W.A.B.
- Frukacz, K. (2019). *Polski reportaż książkowy. Przemiany i adaptacje*. Wydawnictwo Uniwersytetu Śląskiego.
- Górecki, W. (2013). *Abchazja*. Czarne.
- Grabowski, Ł., Hebal-Jezierska, M. (2016). O różnych korpusowych metodach badawczych – próba krytycznej refleksji. *Komunikacja Specjalistyczna*, 11, 65–83.
- Janiszewski, M. (2014). *Dom nad rzeką Loes*. Czarne.
- Kęskrawiec, M. (2010). *Czwarty pożar Teheranu*. W.A.B.
- Kniaginina, M., Pisarek, W. (1966). *Język wiadomości prasowych*. Ośrodek Badań Prasoznawczych RSW „Prasa” w Krakowie.
- Kopczyński, K., Sajewicz, A. (2017). *Dybuk. Opowieść o nieważności świata*. Agora.
- Mańczak, W. (1996). *Problemy językoznawstwa ogólnego*. Ossolineum.
- Maryl, M., Piasecki, M., Walkowiak, T. (2018). Literary exploration machine a web-based application for textual scholars. W: M. Piasecki (red.), *Selected papers from the CLARIN Annual Conference 2017* (s. 128–144). Linköping Electronic Conference Proceedings. <https://doi.org/10.3384/ecp147>.
- Magenowa, M. (1965). Możliwości i niebezpieczeństwa metod matematycznych w poetyce. W: M. Magenowa (red.), *Poetyka i matematyka* (s. 5–20). Instytut Badań Literackich PAN.
- Mikołajewski, J. (2015). *Wielki przyptyw*. Dowody na Istnienie.
- Milewski, J. (2008). *Dym się rozwiewa*. Zys i S-ka.
- Ostrowska, K. (2023). *Reportaż książkowy w świetle językoznawstwa statystycznego. Składnia, styl, gatunek*. Uniwersytet Jana Kochanowskiego w Kielcach.
- Pająchkowska, A. (2019). *Wędrowny zakład fotograficzny*. Czarne.
- Pawlikowska, B. (2006). *Blondynka na Kubie. Na tropach prawdy i Ernesta Che Guevary*. Słowne.
- Pawłowski, A. (2001). *Metody kwantytatywne w sekwencyjnej analizie tekstu*. Uniwersytet Warszawski.
- Piechota, M. (2017). Kiedy poeta pisze reportaż – wywoływanie obrazu w *Wielkim przyptywie* Jarosława Mikołajewskiego. W: I. Hofman i D. Kępa-Figura (red.), *Współczesne media. Gatunki w mediach*, t. 1, *Zagadnienia teoretyczne. Gatunki w mediach drukowanych* (s. 345–361). Uniwersytet Marii Curie-Skłodowskiej.
- Piskała, K. (2014). *Dryland*. W.A.B.
- Rachwałowa, M. (1986). *Słownictwo tekstów naukowych*. Zakład Narodowy im. Ossolińskich.
- Rosiak, D. (2015). *Ziarno i krew. Podróż śladami bliskowschodnich chrześcijan*. Czarne.
- Sambor, J. (1972). *Słowa i liczby. Zagadnienia językoznawstwa statystycznego*. Zakład Narodowy im. Ossolińskich.
- Smoleński, P. (2006). *Izrael już nie frunie*. Czarne.
- Stasiuk, A. (2004). *Jadąc do Babadag*. Czarne.
- Szabłowski, W. (2010). *Zabójca z miasta moreli. Reportaże z Turcji*. Czarne.
- Szczygieł, M. (2006). *Gottland*. Czarne.
- Wanat, E. (2018). *Deutsche nasz. Reportaże berlińskie*. Świat Książki.
- Wlekły, M. (2019). *Raban. O kościele nie z tej ziemi*. Agora.
- Wojtak, M. (2019). *Wprowadzenie do genologii*. Uniwersytet Marii Curie-Skłodowskiej.
- Wolny-Zmorzyński, K. (2004). *Reportaż – jak go napisać? Poradnik dla słuchaczy studiów dziennikarskich*. Wydawnictwa Szkolne i Pedagogiczne.
- Zarębina, M. (1970). Rola wyrazów w słowniku i tekście. *Język Polski*, 50, 33–46.

Katarzyna Ostrowska, Marek Ruszkowski

Zarębina, M. (1971). Najczęstsze wyrazy polszczyzny mówionej. *Język Polski*, 51, 336–347.

Zarębina, M. (1985). *Próba statystycznej analizy słownictwa polszczyzny mówionej (synteza danych liczbowych)*. Zakład Narodowy im. Ossolińskich.

Zasina, A. (2018). Językoznawstwo korpusowe. Empiryczne podejście w badaniach humanistycznych. *Dziennikarstwo i Media*, 9, 169–178.