

Krystyna Płachcińska

O autorstwie "Roksolanek" Zimorowica : w świetle analiz stylometrycznych

Pamiętnik Literacki : czasopismo kwartalne poświęcone historii i krytyce
literatury polskiej 71/2, 165-187

1980

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, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach
dozwolonego użytku.

KRYSTYNA PŁACHCIŃSKA

O AUTORSTWIE „ROKSOLANEK” ZIMOROWICA W ŚWIETLE ANALIZ STYLOMETRYCZNYCH

Znany jest historii literatury nie rozstrzygnięty do dziś spór o autorstwo *Roksolanek*. Przyczynę i dzieje tego sporu wystarczająco jasno i przejrzyście wyłożył Stanisław Adamczewski w swej monografii poświęconej *Sielankom nowym ruskim* Józefa Bartłomieja Zimorowica¹. Zbędne staje się wobec tego referowanie wystąpień kontrpartnerów tego nietypowego dialogu. Dość przypomnieć, że teżą starszą jest rozdzielenie *Roksolanek* i *Sielanek* między Szymona a Bartłomieja Zimorowiców, nowszą zaś przypisywanie autorstwa obu tych dzieł Bartłomiejowi. Kolej zatem tylko na zgrupowanie, usystematyzowanie i syntetyczną krytykę argumentów używanych w sporze, a także — wreszcie — na zastanowienie się, co można powiedzieć na ten temat dzisiaj.

Otóż argumenty, jakimi posługiwali się zwolennicy obu wykluczających się tez, pochodzą z trzech kręgów. Te, które z jednej strony, wobec niepodważalności ich treści faktograficznej, uznać musimy w całości, lecz z drugiej strony ich dowolne interpretowanie przez reprezentantów opozycyjnych stanowisk budzi nasz sprzeciw, to argumenty biograficzno-historyczne.

Pierwszą „kością niezgody” z tego zakresu stał się wiek Szymona Zimorowica. Wiadomo było zawsze, że zmarł on jako bardzo młody człowiek. Byłoby faktem dość dziwnym, że jego talent bez etapu dojrzewania dał tak szybko owoc doskonały, gdyby nie to, iż tłumaczono ten fakt nadzwyczajnymi uzdolnieniami. Młody wiek więc nie tylko nie zaprzeczał autorstwu Szymona, lecz wręcz przeciwnie, wspomógł przeciwstawienie *Roksolanek* jako błysku talentu młodzieńca *Sielankom* jako wytworowi „grubej” poezji starszego brata. Tezę tę mocno podważyło odkrycie, iż młody Zimorowicz zmarł mając lat 21, a nie 26, jak dotąd sądzono². Zmniejszenie prawdopodobieństwa tezy jednak jej nie obaliło. To, co wiadomo na pewno o życiu Szymona³, nie jest w stanie przechylić wagi na żadną ze stron: ani fakt, iż w wieku 19 lat został

¹ S. Adamczewski, *Oblicze poetyckie Bartłomieja Zimorowicza*. Warszawa 1928, s. 4–25.

² Zob. K. Heck, *Kto jest autorem „Roksolanek” pod imieniem Szymona Zimorowicza wydanych*. „Rozprawy AU. Wydział Filologiczny” 1905. I odfitka.

³ Źródłem wiadomości biograficznych jest praca K. Hecka pt. *Materiały do biografii Józefa Bartłomieja i Szymona Zimorowiczów (Ozimków)* (cz. 1. Kraków 1895).

podpiskiem w kancelarii pisarza sądu miejskiego, ani stwierdzenie, iż brał udział w nocnej bójce (co przecież nie zaprzecza tkliwemu i poetycznemu usposobieniu), ani okoliczność, iż ostatni rok życia aż do śmierci spędził w Krakowie, prawdopodobnie na nauce, ani też poświadczony przez Bartłomieja⁴ fakt napisania krótko przed śmiercią utworu zatytułowanego *Roxiades*, w którym wysławiał oblubienicę brata, Katarzynę Duchnicównę — pod imieniem Spirynzyny. Do niezaprzeczalnych faktów należy też zbieżność okresu potwierdzonych zabiegów Bartłomieja o rękę Rozalii Groszawerówny (połączonych z grą na pozytywie i na klawikordzie oraz ze śpiewaniem piosenek) i wydania *Roksolanek* (1654). Usiłowanie nagięcia tych faktów do jednej z hipotez traci jednak zawsze walor dowodowy.

Czy *Roxiades* mogą oznaczać *Roksolanki*? Tak, gdyż zawierają w nazwie ten sam rdzeń *Rox-*, i tu, i tu pojawia się imię Spirynzyny, a także i tu, i tu opiewana jest Duchnicówna. Nie, gdyż „Roxolania”, „roxolański” były słowami używanymi, ale „*Roxiades*” jest wyrazem obcym, nieznanym, neologizmem Zimorowica. Nie, gdyż Duchnicówna w *Ruskich pannach* nie imieniem Spirynzyny, lecz Lilidory jest obdarzona. Poza tym zbieżność czasu zalotów do Groszawerówny z rokiem wydania *Roksolanek* może być przypadkowa. Zupełnie zaś pozbawione realnych podstaw, choć prawdopodobne psychologicznie, są przypuszczenia, że Bartłomiej jako człowiek piastujący urzędy (rajcy i burmistrza lwowskiego) nie chciał, by romansowe piosneczki adresowane do Rozalii z nim łączono i dlatego wydał je pod nazwiskiem brata. Ostatni argument biograficzny — to słowa Stanisława Biezanowskiego, profesora Akademii Krakowskiej i poety panegirycznego, w których (pominąwszy zwykłe hiperbole panegiryczne) czytamy o potrójnych zasługach Bartłomieja Zimorowica na niwie literackiej:

*Mariae seu cantica Lechis e Latio transfert, seu Roxica prata pererrat, seu locat ad Tyram Chodkievi nobile saxum*⁵.

Człon środkowy o przebieganiu roksolańskich niw nie może zaś odnosić się do *Sielanek*, gdyż dziełko Biezanowskiego ukazało się w roku 1659. Nic jednak nie obliguje nas do łączenia z tym sformułowaniem właśnie *Roksolanek*.

Drugą grupę argumentów, jeszcze bardziej zawodnych i interpretowanych dowolnie w obie strony, stanowią wykładnie słów tekstu *Roksolanek* i *Sielanek*. Oto dwa najczęściej przypominane fragmenty:

Semianowi umrzeć nie dadzą sielanki,
Nie dopuszczą zamierchnąć i mnie roksolanki.
(*Kobeznicy*, w. 97–98)

Tobie ma willaneska, robotne sielanki,
Tobie służyć gotowe moje roksolanki;
Tobie się kwoli moja melodyja pieści,
Gdy się ozywa palcem mym w głosach czterdzieści;
(*Zalotnik*, w. 199–202)

⁴ J. B. Zimorowic, *Leopolis triplex*. W: *Pisma do dziejów Lwowa odnoszące się [...]*. Wydał K. Heck. Lwów 1899, s. 200.

⁵ S. Biezanowski, *Acroama panegyricum honori urbis Leopolensis [...] dictatum*. Cracoviae 1659. Inform. i cyt. za: Heck, *Kto jest autorem „Roksolanek” [...]*, s. 352–354.

Oto np. pierwszy z tych fragmentów miał aż trzy różne interpretacje: przeciwstawiającą (aczkolwiek dość opacznie) autora *Sielanek* autorowi *Roksolanek*⁶, drugą — utożsamiającą Semiana z Simonidesem, a *Roksolanki* odnoszącą do autora tych słów, czyli do Bartłomieja⁷, wreszcie interpretację idącą w kierunku konsekwentnej mistyfikacji: skoro Bartłomiej podpisał pod *Sielankami* Szymona, to również jako wypowiedziane przez niego należy odczytać i te wersy (czyli *Sielanki* są dziełem Bartłomieja, a *Ruskie panny* Szymona)⁸. Wykładnia ta dotyczy również fragmentu z *Zalotnika*, a przeciwstawia się jej teza, iż wersy te wypowiada Bartłomiej w swoim imieniu, kładąc na jednym biegunie „moje roksolanki” jako już gotowe, a na drugim „robotne sielanki” jako „tworzące się, pracę kosztujące”⁹.

Podobnie nazwanie Szymona w *Żalobie* „hymnistą gładkobrzmiącej mowy” używane było jako dowód autorstwa poezji doskonałej¹⁰, ale też inaczej — jako dowód posługiwania się mową gładkobrzmiącą, czyli łacińską (tu owe nieznane *Roxiades*)¹¹.

Równie dowolne i nie umotywowane jest wysnuwanie ze słów:

Nie wyprawa bogata, nie posag dostatni,
Ani spadek po krewnych bezpłodnych ostatni,
Ale pieszczona miłość, rozmowy łagodne
Dały do twej niewoli ręce me swobodne;
(*Zalotnik*, w. 149–152)

— wniosku o tym, iż adresatką ich była Rozalia Groszawerówna¹². Fragment zaś:

Jeżeli kiedy moje rymy grube
Będą na potym nowej młodzi lube,
Już nie ja przez nie, lecz ty, ulubiony,
Będziesz na lata potomne wsławiony.
(*Żaloba*, w. 131–134)

— w myśl jednej z hipotez wyjaśniał powód mistyfikacji autorskiej *Sielanek*, upatrywany w chęci złożenia hołdu pamięci brata¹³, a można przecież rozumieć go jako mówiący o wsławianiu Szymona w inny sposób (jak wsławia osobę opłakiwaną tren).

Najważniejszą grupą argumentów są jednak spostrzeżenia wynikłe z analizy językowo-stylistycznej. Jeśli zaś pominiemy tu rozpowszechnione przez czas dłuższy mniemanie o rzekomej rażącej różnicy stylu między *Ruskimi pannami* a *Sielankami*, nie potwierdzone analizą¹⁴, to okaże się, że wymowa tej grupy argumentów popiera

⁶ K. Jarecki, *Kilka uwag o utworach Szymona Zimorowicza*. „Pamiętnik Literacki” 1903, s. 255.

⁷ Heck, *Kto jest autorem „Roksolanek”* [...], s. 355.

⁸ Adamczewski, *op. cit.*, s. 21.

⁹ Heck, *Kto jest autorem „Roksolanek”* [...], s. 359–360.

¹⁰ Jarecki, *op. cit.*, s. 249.

¹¹ Zob. Heck, *Kto jest autorem „Roksolanek”* [...], s. 364.

¹² *Ibidem*, s. 344–345.

¹³ Zob. Cz. Hernas, *Barok*. Warszawa 1973, s. 56–57.

¹⁴ Kształtowanie się tych poglądów przedstawia Adamczewski (*op. cit.*, s. 4–17).

teżę tożsamesgo autorstwa obu dzieł. Po pełny rejestr zbieżności między dwoma interesującymi nas utworami łatwo sięgnąć do pracy tego, który zbieżności te wykrył, stosując metodę drobiazgowej analizy, a którego praca pozostała, nie wiadomo dlaczego, prawie nie zauważona. Mowa oczywiście o Adamczewskim¹⁵. Przypieczmy tylko dla przykładu niektóre z wykrytych zbieżności: rozmaitość epitetów, ta sama świeżość obrazów, te same przenośnie, podobne traktowanie przyrody — z dużym udziałem czynnika estetycznego, spora doza autotematyzmu, przewaga zawodu miłosnego nad uczuciem szczęśliwym, miękkość wyrazu w motywach śmierci, itd. Warto przypomnieć tu jeszcze uwagę językoznawcy o poświadczonym rymami wspólnym dla obu utworów mazurzeniu. Jeśli Simonides mógł wyzbyć się mazuryzmów, czemu miałyby się to nie udać Szymonowi Zimorowicowi, gdyby to on był autorem *Roksolanek*?¹⁶

Czyż więc podejmować ten temat raz jeszcze? Dzieło literackie istnieje przecież niezależnie od tego, jakiemu twórcy je przypiszemy. Ważne jest przecież tylko własne piękno tekstu, jego oddziaływanie estetyczne. Ale dociekliwość filologiczna nie pozwala przejść obok tej sprawy obojętnie, każe przynajmniej próbować ponownego podjęcia sprawy autorstwa interesującego nas dzieła.

Cóż zatem czynić? W sytuacji gdy każdy argument historyczny i historycznoliteracki interpretować można równie dobrze w dwie strony, pozostaje wyniki analizy Adamczewskiego sprawdzić za pomocą badań stylometrycznych. Zastanowić się jednak trzeba na początku, jakie aspekty tekstów badaniom tym warto poddać. Czym powinny różnić się *Roksolanki* od *Sielanek nowych ruskich*, jeśli nie są dziełem jednej ręki? Jakie cechy tych tekstów uznamy hipotetycznie za dostatecznie różnicujące? Mierzyć i liczyć można bowiem wszystko, ale skoro badanie kompleksowe jest bez użycia maszyn matematycznych niemożliwe (a przy tym niepotrzebne), to rzeczą badacza jest zdecydować, co ma być poddane badaniom ilościowym.

W tym oto przypadku, gdy mamy do czynienia z dwoma braćmi, a więc z twórcami wywodzącymi się z tej samej okolicy kraju i z tego samego czasu, ale o różnym podobno stopniu wykształcenia (Bartłomiej pobierał nauki tylko w szkole lwowskiej, a Szymon również w Krakowie), wydaje się, że ich język powinien być zróżnicowany najbardziej w zakresie słownictwa. W związku z tym właśnie zjawiska leksykalne powinny zostać poddane badaniom statystycznym.

Powstaje jednak od razu następne pytanie: jakie teksty można ewentualnie uznać za przeciwstawne sobie? Pozornie problem jest prosty: *Roksolanki* przeciwstawiamy *Sielankom*. W istocie jednak nie możemy porównywać tych utworów traktowanych globalnie, gdyż nie jesteśmy przekonani o ich jednolitości pod względem autorstwa (to, że tworzą całości wydawnicze, nie jest bowiem żadnym argumentem). Zarówno *Roksolanki*, jak i *Sielanki* mogą przecież składać się z fragmentów pisanych przez Szymona przeplatanych fragmentami autorstwa Bartłomieja. O ile w obrębie *Sielanek* wydzielenie tych utworów, których z pewnością nie mógł napisać Szymon, jest łatwe (przez porównanie faktów, o których mowa w tych sielankach, z datą śmier-

¹⁵ *Ibidem*, s. 148–171.

¹⁶ Zob. K. Nitsch, *Z historii polskich rymów*. Warszawa 1912, s. 34–35.

ci Szymona), o tyle żadne wiadomości biograficzne ani historyczne nie pozwolą wydzielić pieśni *Roksolanek*, których nie mógłby napisać Bartłomiej. Żywotny jest więc problem jednolitości autorskiej *Ruskich panien*.

Badanie jednolitości „Roksolanek”

Pierwszą czynnością było sporządzenie pełnego słownika frekwencyjnego tego utworu, co pozwoliło ustalić rozkład częstości wyrazów. Tabela zawierająca ów rozkład zamieszczona jest na s. 170—171.

Na wyniki tej tabeli składa się jednak 71 pieśni, między którymi należy przeprowadzić linię podziału, jeśli taki istnieje i jeśli jest to możliwe. Przeciętna długość pieśni wynosi 152 wyrazy, tylko pieśń dziewczosłeba jest odcinkiem długim ($N = 2685$ wyrazów). Odpadają zatem wszelkie badania mogące doprowadzić do podziału według częstości względnej użycia jakiegoś wyrazu. Są to bowiem częstości bardzo małe, a więc wymagające prób o dużej liczebności. Dla przykładu wyraz „ja” posiada w *Roksolankach* częstość względną

$$p_{\text{emp.}} = \frac{330}{13260} = 0,0249$$

Jeśli przyjmiemy dokładność oszacowania $\delta = 0,3$, a żadaną minimalną liczebność (N) próby obliczymy przekształcając wzór określający zależność funkcyjną między δ , N i p ¹⁷:

$$N = \frac{t_{\alpha}^2}{\delta^2 p}$$

gdzie t — wartość odczytana z tablic rozkładu Gaussa dla przyjętego z góry współczynnika ufności α (zakładamy możliwość pomyłki w 5 przypadkach na 100, a więc $\alpha = 0,05$, a odczytane dla 0,05 $t_{\alpha} = 1,96$)

N — liczebność próby

p — częstość względna danego elementu

— to okaże się, że w przypadku wyrazu „ja” konieczna jest próba o wielkości przynajmniej

$$N = \frac{(1,96)^2}{(0,3)^2 \cdot 0,0249} = 1714 \text{ wyrazów}$$

Wobec wyrazów rzadszych, np. najliczniej występującego wyrazu pełnoznacznego, jakim jest w *Roksolankach* „serce”, konieczna jest jeszcze większa próba. Stosowane wobec niego wnioskowanie statystyczne miałoby sens tylko wtedy, gdybyśmy zapewnili sobie próbę o wielkości przynajmniej

$$N = \frac{(1,96)^2}{(0,3)^2 \cdot 0,0069} = 6186 \text{ wyrazów}$$

¹⁷ Wzór zaczerpnięty z pracy J. Sambor pt. *Słowa i liczby. Zagadnienia językoznawstwa statystycznego* (Wrocław 1972, s. 263). Wszelkie pozostałe wykorzystane wzory, jeśli nie podano ich źródła, pochodzą z tej podstawowej polskiej pracy poświęconej zreferowaniu zastosowania metod statystycznych w językoznawstwie i w stylistyce.

Ranga	Wyraz	Liczba wyrazów (w_i)	Częstość (f_i)
1	ja	1	330
2	nie	1	328
3	w	1	307
4	z	1	285
5	on	1	263
6	i	1	240
7	ty	1	212
8	na	1	203
9	mój	1	154
10	ten	1	146
11	być	1	130
12	twój	1	119
13	który	1	114
14	a	1	108
15	do	1	97
16	że	1	94
17	serce	1	91
18	swój	1	82
19	za	1	76
20	tak	1	72
21	mieć	1	68
	jako to	2	67
	gdy oko	2	64
	ani wszytek	2	56
	już się świat	3	47
	wy	1	43
	co miły	2	42
	chcieć dać	2	39
	ale ogień sam	3	38
	przez	1	35
	kiedy lecz nad po	4	33

Wyraz	Liczba wyrazów (w_i)	Częstość (f_i)
dla tylko	2	32
ciało kto	2	31
u	1	29
móc o	2	28
czas Kupido miłość piękny rok	5	27
człowiek wasz wiedzieć	3	25
jeden od przyjaciół różany	4	24
jeśli kiedy kochany też	4	23
bez przecię	2	22
bo młody płomień twarz uroda	5	21
barzo czoło głowa nadobny żeby	5	20
choć choć dusza dwa dzień gdzie ode skoro wiele	9	19
tam	1	18
bowiem brać	10	17

Wyraz	Liczba wyrazów (w_i)	Częstość (f_i)	Wyraz	Liczba wyrazów (w_i)	Częstość (f_i)
czemu				14	15
dobry				14	14
noc				7	13
przed				13	12
przy				18	11
przyjąć				20	10
teraz				28	9
żyć				39	8
abo	13	16		43	7
drugi				70	6
dziewczyna				89	5
Halina				195	4
innszy				240	3
-li				544	2
ręka				1984	1
skrzydło					
strzała					
także					
tu					
widzieć					
wziąć					

$$N = 13260$$

$$W = 3429$$

Pieśni *Roksolanek* mają zaś od 85 do 334 wyrazów, a więc podziału ze względu na częstość użycia danych wyrazów, nawet najpowszechniejszych, nie sposób przeprowadzić. Dopiero częstości względne rzędu $0,3 \div 0,5$ są w stanie zapewnić żądaną dokładność. W częstościach takich mogą jednak występować tylko duże grupy wyrazów, np. poszczególne części mowy. Możliwe zaś, że linia podziału przeprowadzona na tej podstawie dałaby obraz fałszywy. Różnice bowiem w częstościach użycia czasowników czy przymiotników, choć statystycznie istotne, mogłyby być spowodowane nie szczególną predylekcją autora do stosowania danej części mowy, lecz po prostu tematem, który raz wymagał dużej liczby czasowników, a raz zaimków. Również więc przy tej metodzie konieczna byłaby znacznie liczniejsza próba.

Zwiększyć zaś liczebność próby możemy jedynie przechodząc od leksemów do fonemów. Jeśli bowiem na pieśń składa się przeciętnie około 150 wyrazów, to fonemów mniej więcej 1000—1500, a taka liczebność pozwala już na wysnuwanie wniosków statystycznych (przy fonemach najczęstszych z dużą dokładnością, przy rzadszych natomiast z większym błędem, $\delta = 0,4 \div 0,5$). Tymczasem jednak należałoby zaobserwować wstępnie jakieś prawidłowości, na których podstawie można by sformułować hipotezy dla poddania ich weryfikacji statystycznej, a jest to na terenie fonologii trudne. Nie wiadomo więc, co obserwować, nie wiadomo, gdzie może kryć się róż-

nica między ręką Bartłomieja a Szymona. Trudno przecież obserwować i obliczać wszystko: samogłoski i spółgłoski, poszczególne fonemy i ich bezpośrednie następstwa po sobie. Jest to zajęcie pracochłonne, a najpewniej byłoby bezowocne, jako że elementy systemu fonologicznego języka dwóch braci nie powinny różnić się pomiędzy sobą znacznie.

Pozostawmy więc w sferze słownictwa, gdzie zjawiska są najłatwiej zauważalne i możliwe do poddania ich badaniom kwantytatywnym.

Spróbujmy powtórzyć tu interesujące obliczenia, jakie wobec 10 listów św. Pawła zastosował G. Herdan¹⁸. Dociekał on stopnia filiacji słownictwa owych listów podzielonych na 9 grup, pragnąc wyodrębnić listy szczególnie sobie bliskie pod względem leksyki. Stosując metodę Herdana do analizy pokrewieństwa poszczególnych pieśni *Roksolanek* zgromadzono w tabeli następujące dane¹⁹:

— liczbę wyrazów danej pieśni, które nie występują już po raz drugi w całym zbiorze (tzw. *hapaxlegomena*),

— liczbę wyrazów w pieśni, które powtarzają się w 1 z pozostałych pieśni,

— liczbę wyrazów powtarzających się w 2 innych pieśniach, itd. aż do 9 pieśni.

Gromadzenie dalszych danych mija się z celem, gdyż słowa wspólne dużej liczbie pieśni to wyrazy gramatyczne, które muszą występować we wszystkich tekstach i nie tworzą obrazu zróżnicowania leksykalnego. Najistotniejsze w tej analizie są wyrazy wspólne dwóm, trzem i czterem tekstom, czyli wyrazy rzadkie. Oto fragment tabeli:

	<i>Hpx</i>	Liczby wyrazów wspólnych z 1, 2, ..., 9 pieśniami									Razem
		1	2	3	4	5	6	7	8	9	
Ded.	60	16	4	7	7	5	3	1	1	2	106
Dz.	929	163	77	71	32	22	13	11	13	12	1343
I, 1	52	12	11	9	1	4	4	1	2	3	99
I, 2	44	13	9	6	2	1	1	2	3	1	82
I, 4	42	9	1	9	5	2	—	3	1	3	75
I, 5	25	2	3	6	1	2	—	1	2	1	43
I, 7	47	5	11	6	3	2	6	2	3	4	89
I, 9	51	15	6	9	4	4	3	2	1	—	95
II, 14	57	19	10	8	7	3	1	3	4	3	115
II, 19	108	15	8	14	6	5	2	3	4	2	167
III, 5	17	3	4	1	1	3	1	1	—	2	33
III, 7	76	23	8	6	5	4	7	1	1	1	132
III, 19	77	16	7	11	5	7	4	1	3	3	134
III, 20	35	4	4	3	2	—	—	1	2	2	53

A oto przykład badania wzajemnej zależności pieśni dedykacyjnej z roksolanką 14 chóru II:

¹⁸ G. Herdan, *Type-Token-Mathematics*. Mouton 1960, s. 104—108. Informacja o metodzie Herdana zaczerpnięta z pracy Sambor (*op. cit.*, s. 114—118).

¹⁹ Dane te zgromadzone zostały na podstawie drugiej połowy słownika *Roksolanek*, od N do Ż.

			H_{px}	Liczby wyrazów wspólnych z 1, 2, ..., 9 pieśniami									Razem (b_j)
				1	2	3	4	5	6	7	8	9	
Częstości	Empiryczne	Ded.	60	16	4	7	7	5	3	1	1	2	106
		II, 14	57	19	10	8	7	3	1	3	4	3	115
		Razem (a_i)	117	35	14	15	14	8	4	4	5	5	221 = N
	Teore- tyczne	Ded.	56	17	7	7	7	4	2	2	2	2	
		II, 14	61	8	7	8	7	4	2	2	3	3	

Dane zawarte w części „Częstości empiryczne” pochodzą z tabeli poprzedniej a więc wprost z obliczeń dokonanych na utworze. Dane w części „Częstości teoretyczne” zostały obliczone (na podstawie twierdzenia, że prawdopodobieństwo iloczynu zajścia zdarzeń niezależnych równe jest iloczynowi prawdopodobieństw poszczególnych zdarzeń) ze wzoru:

$$f'_{ij} = \frac{a_i b_j}{N}$$

gdzie: f'_{ij} — teoretyczna częstość wystąpienia w danej pieśni (j) danej liczby wyrazów wspólnych z i pieśniami

$$i = 1, \dots, 10$$

$$j = 1, 2$$

Częstości teoretyczne są więc iloczynami poszczególnych pozycji zawartych w wierszu i w kolumnie „Razem”, podzielonymi przez N . Gdyby porównywane pieśni były pod względem słownictwa zróżnicowane, to liczby wyrazów wspólnych też różniłyby się od siebie w sposób nieprzypadkowy. Sprawdzenie istotności różnic występujących między liczbami faktycznymi a tymi, których oczekivalibyśmy teoretycznie, dokonujemy, podobnie jak Herdan, stosując test χ^2 :

$$\chi^2 = \sum_{i=1}^{10} \sum_{j=1}^2 \frac{(f_{ij} - f'_{ij})^2}{f'_{ij}}$$

gdzie: f_{ij} — częstości empiryczne

f'_{ij} — częstości teoretyczne

W naszym przypadku liczba stopni swobody wynosi 9, a więc przyjmując współczynnik ufności 0,01 odczytujemy z tablic rozkładu χ^2 teoretyczną wartość sprawdzianu χ^2_a . Wynosi ona 21,666. Tymczasem obliczona według powyższego wzoru empiryczna wartość χ^2 dla *Dedykacji* i pieśni II, 14 wynosi 6,75. Ponieważ jest to wartość niższa od teoretycznej, nic nie skłania nas do odrzucenia hipotezy o braku istotnych różnic leksykalnych między tymi fragmentami. Podobne przebadanie stosunku innych pieśni do *Dedykacji*, a wrywkowo także innych pieśni między sobą, też nie wykazało znacznych z punktu widzenia rachunku prawdopodobieństwa różnic. Empiryczne wartości sprawdzianu χ^2 wahają się między 2,38 a 19,75.

Roksolanki wyglądają więc na blok jednolity pod względem słownictwa, przynajmniej zaś nic nie wskazuje na wewnętrzne zróżnicowanie. Taka ostrożność sformu-

lowania jest konieczna, ponieważ wnioski negatywne są pewne, a pozytywne jedynie bardzo prawdopodobne. Jeśli nawet w *Ruskich pannach* widać ślady ręki więcej niż jednego autora, to są one bardzo nieliczne i ułamkowe, a główny twórca wkomponował je w tok swoich myśli.

Ustaliliśmy więc, co prawda nie z pewnością, ale z dużym prawdopodobieństwem, że tekstem *A*, który zestawiać będziemy z utworami Bartłomieja, jest traktowany globalnie utwór *Roksolanki*, składający się z 13 260 wyrazów, w tym 3429 haseł.

Tekstowi *A* przeciwstawimy tekst *B*, wybrany spośród tych sielanek, których autorem żadną miarą nie mógł być Szymon. Są to:

— *Roczyzna*, ponieważ traktuje o śmierci Szymona i datowana jest przez Bartłomieja na rok 1647,

— *Żaloba*, gdyż również jej tematem jest śmierć Szymona,

— *Kozaczyzna* i *Burda ruska*, ponieważ dotyczą wydarzeń historycznych późniejszych od śmierci Szymona o 19 lat,

— *Filoreta*, jako że opiewa śmierć pierwszej żony Bartłomieja, zmarłej w r. 1653, a datowana jest przez poetę jako pisana 2 lata po tym fakcie.

Co do pozostałych sielanek, o których twierdzi się, że powstały po r. 1629²⁰, można zawsze wysunąć zastrzeżenia, np. w *Trużenikach* pojawia się wzmianka o śmierci czworga dzieci poety. Powstać mogła najwcześniej w r. 1639, ale fragment ten mógł zostać w ostatecznej redakcji włączony do sielanki napisanej dawniej; ton religijny *Przenosin* nie świadczy jeszcze, że powstały „nie wcześniej jak w okresie tworzenia hymnów religijnych i pobożnego nastroju poety”; podobnie w *Zjawieniu* nie dowodzi późnego czasu powstania wspomnienie „o swobodzie i pokusach młodości”²¹; także w *Winiarzach* nie ma fragmentu, który by kazał twierdzić z pewnością, że sielanka ta napisana została dopiero po nabyciu przez poetę winnicy (co nastąpiło w r. 1644).

Zatem na tekst *B*, przeciwstawiany przez nas *Roksolankom* (*A*), składa się 5 sielanek. Analizowanie ich wszystkich byłoby niepotrzebną stratą czasu, gdyż próba tak liczna nie jest konieczna dla zachowania pożądanej dokładności 30%. Tak więc dla badań najbardziej pracochłonnych, dla sporządzenia słownika frekwencyjnego i rozkładu częstości wyrazów wybrać należało część z owych 5 utworów. Wybór padł na *Żalobę* (*B*₁), której styl podobny jest do stylu *Roksolanek*, i na sielankę z gruntu inną, nie budzącą wątpliwości, iż spod ręki Szymona nie pochodzi (zarówno ze względu na temat osnuty na tle wydarzeń zaszłych po jego śmierci, jak i na rzekomo całkowicie odmienny typ artyzmu), na *Burdę ruską* (*B*₂).

Porównanie parametrów częstości rozkładu wyrazów

Rozkład częstości wyrazów *Żaloby* zawiera tabela na s. 175.

²⁰ Referuje te poglądy J. Łoś w *Objaśnieniach* do swej edycji *Sielanek nowych ruskich* (Kraków 1916, s. 159–171).

²¹ *Ibidem*, s. 166, 169.

Wyraz	Liczba wyrazów (w_i)	Częstość (f_i)
w	1	51
z	1	39
nie	1	35
on	1	34
na	2	24
ty		
i	2	23
twój		
być	1	18
a	1	15
ja	2	13
swój		
już	4	12
który		
mój		
żał		
do	2	11
ten		
czas	1	10
kiedy	4	9
nad		
o		
tu		
co	7	8
dzień		
grób		
my		
po		
Tymorynna		
wy		
dla	9	7
Hallicynia		

Wyraz	Liczba wyrazów (w_i)	Częstość (f_i)
Helenora		
jako		
łza		
oko		
Pneumancya		
świat		
że		
ale	11	6
dać		
Licydyna		
mieć		
półki		
rok		
serce		
się		
siostra		
śmierć		
wszytek		
dusza	10	5
gdy		
kędy		
niebo		
nowy		
teraz		
to		
usta		
wasz		
za		
	27	4
	45	3
	126	2
	650	1

$$N = 1808$$

$$W = 908$$

Tekst *Żaloby* składa się więc z $N = 1808$ wyrazów, w tym z $W = 908$ wyrazów różnych, czyli że jego słownik zawiera 908 wyrazów. Na podstawie tego rozkładu częstości obliczamy główne parametry statystyczne.

- Średnia częstość $\bar{f}$ występowania wyrazu i odchylenie średnie s
Wskaźnik $\bar{f}$ obliczamy posługując się znanym wzorem:

$$\bar{f} = \frac{N}{W} = \frac{1808}{908} = 1,991$$

Obliczamy też odchylenie średnie:

$$s = \sqrt{\frac{\sum_{i=1}^N (f_i - \bar{f})^2}{N}} = 3,44$$

i współczynnik zmienności v :

$$v = \frac{s}{\bar{f}} = 1,73$$

Wskaźnik $\bar{f}$ ma tu wartość bardzo niską, a przy tym mały jest też stopień zmienności tej cechy. Taki typ rozkładu częstości świadczy o dominowaniu w tekście wyrazów rzadkich, o małych częstościach, przy jednoczesnym niewielkim zróżnicowaniu częstości.

Dla porównania: w jednej z najkrótszych ksiąg *Pana Tadeusza*, w księdze VI (która jednak i tak jest ponad dwukrotnie dłuższa od tekstu *Żaloby*, wskutek czego porównanie nie jest w pełni miarodajne), $\bar{f} = 2,39$ przy $v = 2,57$. Dla połowy tekstu poematu Mickiewicza (księgi I–VI) $\bar{f} = 5,02$ oraz $v = 5,57$. Te ostatnie wartości są ze względu na równą wielkość próby w pełni porównywalne z odpowiednimi parametrami dla *Listów św. Pawła*, dla których $\bar{f} = 25,27$ przy $v = 12,70$ ²². Na tle tych danych ocenić można, jak dużo jest w *Żalobie* wyrazów rzadkich, a więc jak bogate jest słownictwo tej sielanki. Jego bogactwo jest mniej więcej rzędu bogactwa leksyki *Pana Tadeusza*. Prawie dwukrotnie mniejsza wartość $\bar{f}$ dla *Żaloby* niż dla tekstu Mickiewicza nie świadczy o większym bogactwie słownictwa w niej, skoro *Żaloba* jest znacznie krótsza od porównywanej księgi VI, a wskaźnik $\bar{f}$ zależny jest od wielkości próby. Liczby te przytoczone zostały jedynie dla uzmysłowienia, jak duża jest wartość $\bar{f} = 1,991$ przy $v = 1,73$.

2. Wskaźnik oryginalności słownictwa ($I_{\text{oryg.}}$) podany przez Guirauda²³, mierzący udział hapaxlegomenów w słowniku tekstu

$$I_{\text{oryg.}} = \frac{W_1}{W}$$

gdzie: W_1 — liczba wyrazów o częstości 1,

W — wielkość słownika tekstu

$$I_{\text{oryg.}} = \frac{650}{908} \cdot 100\% = 71,6\%$$

²² Dane zaczerpnięte z pracy J. Sambor pt. *Badania statystyczne nad słownictwem (na materiale „Pana Tadeusza”)* (Wrocław 1969, s. 157). Wszelkie inne dane dotyczące słownictwa *Pana Tadeusza* i *Nowego Testamentu* pochodzą również z tego źródła.

²³ P. Guiraud, *Les Caractères statistiques du vocabulaire*, Paris 1954, s. 34. Cyt. za: Sambor, *Słowa i liczby*, s. 222.

Udział wyrazów o częstościach 1, 2 i 3 łącznie wynosi:

$$\frac{\sum_{i=1}^3 W_i}{W} \cdot 100\% = \frac{821}{908} \cdot 100\% = 90,4\%$$

Dla porównania (z tymi samymi zastrzeżeniami co przy wskaźniku $\bar{f}$, gdyż $I_{\text{orig.}}$ też zależy jest od wielkości próby):

$\frac{W_1}{W} \cdot 100\%$	dla ks. VI <i>Pana Tadeusza</i> wynosi 68,44%
	dla ks. I–VI „ „ „ 53,47%
	dla <i>Listów św. Pawła</i> „ 43,05%
$\frac{\sum_{i=1}^3 W_i}{W} \cdot 100\%$	dla ks. VI <i>Pana Tadeusza</i> wynosi 87,82%
	dla ks. I–VI „ „ „ 78,10%
	dla <i>Listów św. Pawła</i> „ 68,43%

3. Wskaźnik stereotypowości słownictwa ($I_{\text{stereot.}}$) mierzący przeciętną długość odstępów między kolejnymi użyciami wyrazów powtarzających się

Obliczamy go, eliminując z tekstu wszystkie wyrazy nie powtarzające się (*hapax-legomena*), przez porównanie długości tekstu z objętością słownika:

$$I_{\text{stereot.}} = \frac{N - W_1}{W - W_1} = \frac{1808 - 650}{908 - 650} = 4,49$$

Im większa jest wartość tego wskaźnika, tym większe zróżnicowanie ilościowe i bogactwo słownictwa tekstu.

4. Miary różnorodności Gooda (c_n)²⁴, wyznaczające prawdopodobieństwo natrafienia przy n -krotnym losowaniu z tekstu na ten sam wyraz

Jest to jeden z najlepszych wskaźników bogactwa leksykalnego²⁵, obliczany jest bowiem na podstawie rozkładu częstości wyrazów w tekście, a więc posiada walory wskaźnika nieobciążonego, niezależnego od długości badanego tekstu. Ma to w naszym przypadku duże znaczenie, gdyż porównywać będziemy próby o wyrażnie dużej różnicy liczebności. Nie znajdują w związku z tym zastosowania wskaźniki obliczane syntetycznie na podstawie stosunku długości tekstu (N) do liczby występujących w nim różnych wyrazów (W). Mowa o wskaźnikach Paula Guirauda ($W/\sqrt{2N}$) i W. Kuraszkiewicza ($W/\sqrt{N}$)²⁶. Jeśli więc zostanie podany wskaźnik Kuraszkiewicza, to pamiętając o jego obciążeniu należy traktować go tylko jako zastosowany pomocniczo, dla zobrazowania pewnego pobocznego zjawiska.

²⁴ I. J. Good, *The Population Frequencies of Species and the Estimation of Population Parameters*, „Biometrika” 1953.

²⁵ Zob. J. Woronczak, *Metody obliczania wskaźników bogactwa słownikowego tekstów*. W zbiorze: *Poetyka i matematyka*. Warszawa 1965, s. 145.

²⁶ Zob. Sambor, *Badania statystyczne nad słownictwem*, s. 112.

Oto więc podstawowy wskaźnik bogactwa leksykalnego, miary różnorodności Gooda:

— dla dwukrotnego losowania (c_2)

$$c_2 = \frac{\sum f_i^2 W_i - N}{N(N-1)} = \frac{14368 - 1808}{1808 \cdot 1807} = \frac{12560}{3267056} = 0,0038444$$

— dla trzykrotnego losowania (c_3)

$$c_3 = \frac{\sum f_i^3 W_i - 3 \sum f_i^2 W_i + 2N}{N(N-1)(N-2)} = \frac{368120 - 43104 + 3616}{1808 \cdot 1807 \cdot 1806} =$$

$$= \frac{328632}{5900303100} = 0,000055697$$

Rozkład częstości wyrazów w sielance *Burda ruska* zamieszczony jest na s. 179. A oto obliczone dla tego rozkładu cztery parametry:

$$a) \quad \bar{f} = \frac{2356}{1225} = 2,094$$

$$s = 4,79$$

$$b) \quad I_{\text{oryg.}} = \frac{W_1}{W} \cdot 100\% = \frac{830}{1225} \cdot 100\% = 73,8\%$$

$$\frac{\sum_{i=1}^3 W_i}{W} \cdot 100\% = \frac{830 + 132 + 54}{1125} \cdot 100\% = 90,3\%$$

$$c) \quad I_{\text{stereot.}} = \frac{N - W_1}{W - W_1} = \frac{2356 - 830}{1125 - 830} = 5,17$$

d) miary Gooda

$$c_2 = \frac{30712 - 2356}{5548380} = 0,00511$$

$$c_3 = \frac{1400824 - 92136 + 4712}{13060886000} = 0,00010056$$

Obliczone powyżej cztery parametry wyznaczają do pewnego stopnia strukturę leksyki w próbie B_1 i B_2 . Następnym punktem postępowania było obliczenie tychże parametrów dla tekstu *Ruskich panien*. Ponieważ jednak trzy pierwsze z nich wymagają prób o podobnej liczebności, konieczne było wyodrębnienie z tekstu *Roksolanek*, czyli z próby A , podgrupy A_1 odpowiadającej liczebnie próbom B_1 i B_2 . Wybór padł na pieśń dziewczosłęba, gdyż tworzy ona tekst spójny, podobnie jak każda z sielanek. Wskaźnik czwarty, miary Gooda, można już było obliczyć dla całości *Roksolanek*. Porównawczo jednak obliczono go również dla tekstu A_1 .

Wyraz	Liczba wyrazów (w_i)	Częstość (f_i)
z	1	74
on	1	67
i	1	61
na	1	51
w	1	48
nie	1	44
ten	1	30
a	1	28
do	1	24
który	2	21
że		
miasto	2	18
tak		
być	1	17
swój	1	16
mieć	2	15
wszytek		
co	2	14
sam		
jako	1	12
dzień	3	11
już		
za		
ja	3	10
się		
to		
Dorosz	2	9
od		
człowiek	4	8
dać		
gdy		
tylko		
głowa	8	7
kiedy		
musieć		
o		
przedmiejski		
tam		
też		
wał		

Wyraz	Liczba wyrazów (w_i)	Częstość (f_i)
dom	10	6
gdzie		
mieszczanin		
ogień		
pod		
ręka		
straszny		
śmierć		
taki		
żaden		
ale	27	5
bez		
bowiem		
Bóg		
cały		
ciało		
czas		
dwa		
jeden		
jeszcze		
każdy		
kędy		
kto		
miejski		
my		
nasz		
noc		
potym		
prędko		
przed		
skoro		
srogi		
stać		
Tatar		
tatarski		
tu		
wielki		
	32	4
	54	3
	132	2
	830	1

 $N = 2356$
 $W = 1125$

Oto tabela zawierająca rozkład częstości wyrazów pieśni dziewosłęba:

w_i	f_i	w_i	f	w_i	f_i	w_i	f_i	w	f_i
1	76	1	44	2	16	1	9	71	3
1	68	1	40	1	15	3	8	184	2
1	63	1	30	2	14	6	7	991	1
1	62	1	25	1	13	13	6	$N = 2665$ $W = 1342$	
1	49	2	18	2	11	21	5		
1	48	2	17	5	10	25	4		

A oto owe cztery parametry obliczone dla próby A_1 (parametr czwarty również dla próby A):

a) średnia częstość i odchylenie średnie

$$\bar{f} = \frac{2665}{1342} = 1,985$$

$$s = 4,8$$

$$v = 2,41$$

b) wskaźnik oryginalności słownictwa

$$\text{— udział hapaxlegomenów } \frac{991}{1342} \cdot 100\% = 73,8\%$$

$$\text{— udział hapax-, dis- i trislegomenów łącznie } \frac{1246}{1342} \cdot 100\% = 92,8\%$$

c) wskaźnik stereotypowości

$$I_{\text{stereot.}} = \frac{2665-991}{1342-991} = 4,7$$

d) miary różnorodności Gooda
dla próby A_1 :

$$c_2 = \frac{35569-2665}{2665 \cdot 2664} = 0,0046346$$

$$c_3 = \frac{1725973-106707+5330}{2665 \cdot 2664 \cdot 2663} = 0,0000859$$

dla próby A :

$$c_2 = \frac{868334-13260}{13260 \cdot 13259} = 0,0048635$$

$$c_3 = \frac{199219930-2605002+26520}{13260 \cdot 13259 \cdot 13258} = 0,0000843$$

Zestawmy w tabeli wskaźniki charakterystyczne dla prób A , A_1 , B_1 i B_2 :

Parametr		<i>Dziewosłab</i> $A_1 N=2665$	<i>Roksolanki</i> $A N=13260$	<i>Żaloba</i> $B_1 N=1808$	<i>Burda ruska</i> $B_2 N=2356$
średnia częstość i odchylenie średnie	$\bar{f}$	1,985		1,991	2,094
	s	4,8		3,44	4,79
wskaźnik oryginalności słownictwa	$\frac{W_1}{W}$	73,8%		71,6%	73,8%
	$\frac{\sum_{i=1}^3 W_i}{W}$	92,8%		90,4%	90,3%
	$\frac{W}{W}$				
wskaźnik stereotypowości słownictwa	$\frac{N-W_1}{W-W_1}$	4,7		4,49	5,17
miary różnorodności Gooda	c_2	0,0046346	0,0048635	0,0038444	0,00511
	c_3	0,0000859	0,0000843	0,000055697	0,00010056

Widać z tego zestawienia, że wartości w kolumnie A (A_1) nie są identyczne z odpowiadającymi im wartościami w kolumnach B_1 i B_2 . Należy teraz postawić sobie pytanie, czy różnice istniejące między tymi wartościami są z punktu widzenia rachunku prawdopodobieństwa istotne, czy też są nieznaczne, wynikię z przypadku. Otrzymamy tym samym odpowiedź, czy *Dziewosłab* i *Żaloba*, a potem *Dziewosłab* i *Burda ruska* wykazują różnice w strukturze ilościowej słownictwa, czy też różnicy tej nie obserwujemy. Zgodnie ze stałym modelem postępowania w takim przypadku konstruujemy hipotezę zerową o braku istotnych różnic między próbą A_1 i B_1 kolejno dla poszczególnych parametrów i poddajemy następnie tę hipotezę weryfikacji przy pomocy testu istotności.

1. Różnica między średnimi częstościami ($\bar{f}_{A_1} - \bar{f}_{B_1}$)

Ustalając poziom istotności $\alpha = 0,05$ (tzn. uznając za zdarzenie bardzo rzadkie takie, które zachodzi 5 razy na 100) odczytujemy z tablic rozkładu Gaussa odpowiadającą temu zdarzeniu wartość sprawdzianu teoretycznego $t_\alpha = 1,96$. Obliczamy następnie wartość sprawdzianu empirycznego $t_{\text{emp.}}$.

$$t_{\text{emp.}} = \frac{(\bar{f}_{A_1} - \bar{f}_{B_1}) - (\mu_{A_1} - \mu_{B_1})}{\sqrt{\frac{s_{A_1}^2}{N_{A_1}} + \frac{s_{B_1}^2}{N_{B_1}}}}$$

Ponieważ hipoteza zakłada, że średnia wartość częstości średniej: μ_{A_1} i μ_{B_1} — w badanych populacjach są równe, dlatego odjemnik w liczniku ułamka pomijamy.

$$t_{\text{emp.}} = \frac{-0,006}{\sqrt{\frac{23,04}{2665} + \frac{11,83}{1808}}} = -0,049$$

Ponieważ $|t_{\text{emp.}}| < t_a$, wynik nie przeczy hipotezie. Nic zatem nie skłania nas do jej odrzucenia. Nie stwierdziliśmy różnicy między średnimi częstościami pojawienia się wyrazu w obu tekstach.

2. Różnica między wskaźnikami oryginalności słownictwa

Obliczamy wartość sprawdzianu empirycznego dla różnicy między wskaźnikami dotyczącymi udziału „jedynek” (wyrazów o częstościach 1) w tekście, zamieniając liczby podane w procentach na wartości prawdopodobieństw:

w A_1 : 73,8%, czyli $p = 0,738$

w B_1 : 71,6%, czyli $p = 0,716$

$$t_{\text{emp.}} = \frac{p_{A_1} - p_{B_1}}{\sqrt{\frac{(1-p_{A_1})p_{A_1}}{N_{A_1}} + \frac{(1-p_{B_1})p_{B_1}}{N_{B_1}}}} = \frac{0,022}{\sqrt{0,0000725 + 0,0001124}} = 1,62$$

Również więc dla tego wskaźnika zachodzi relacja $|t_{\text{emp.}}| < t_a$, a więc hipotezy odrzucić nie możemy.

Sprawdzian zastosowany wobec wyrazów o częstościach 1 ÷ 3 wygląda następująco:

w A_1 : 92,8%, czyli $p = 0,928$

w B_1 : 90,4%, czyli $p = 0,904$

$$t_{\text{emp.}} = \frac{0,024}{\sqrt{\frac{(1-0,928) \cdot 0,928}{2665} + \frac{(1-0,904) \cdot 0,904}{1808}}} = 2,81$$

W tym przypadku $|t_{\text{emp.}}| > 1,96$, co każe odrzucić hipotezę o braku istotnych różnic między *Żalobą* a *Dziewosłębem* w zakresie prawdopodobieństwa pojawienia się wyrazów o trzech najmniejszych częstościach. W *Dziewosłębie* jest ich z punktu widzenia statystyki o tyle więcej niż w *Żalobie*, że z dużym prawdopodobieństwem wykluczyć możemy działanie przypadku. Teksty te różnią się zatem ze względu na tę cechę. Różnica ta jednak nie jest duża, a wzięwszy pod uwagę istniejącą, mimo starań przeciwdziałających owemu zjawisku, nierówną liczebność prób, możemy temu właśnie, a nie faktycznej różnicy w rozkładzie częstości słownictwa, przypisać wpływ na wartość sprawdzianu większą niż dopuszczalna. Należy przy tym pamiętać, że przy weryfikacji hipotezy o różnicy udziałów ($p_1 - p_2$) test t jest zbyt czuły dla dużych N_1 i N_2 . Mianownik bowiem jest bardzo mały, a więc przy nieznacznych nawet różnicach $p_1 - p_2$ duży jest iloraz $t_{\text{emp.}}$. Jest to podstawowa niedogodność tego testu. Na karb jego zbyt dużej czułości położyć możemy większą niż dopuszczalna (1,96) wartość $t_{\text{emp.}}$ dla wyrazów o częstościach 1 ÷ 3. Przy tym podkreślić należy, że mimo tak znacznej czułości testu wartość $t_{\text{emp.}}$ dla hapaxlegomenów jest niższa od 1,96. Jest to więc argument bardzo ważki.

3. Różnica między średnimi odległościami wyrazów powtarzających się

Różnicy tej nie uda się nam poddać testowi istotności z powodu braku jednego z koniecznych do tego danych: odchylenia średniego odległości wyrazów. Jego obli-

czenie jest zbyt pracochłonne, by warto było tego dokonywać. Różnicę tę wystarczy ocenić intuicyjnie²⁷. Jest nieznaczna.

4. Miary Gooda

Również nie potrafimy poddać tego wskaźnika testowi istotności. Miary te mają bowiem dość skomplikowane rozkłady, nie opracowane jeszcze przez teorię statystyki. W związku z tym nie znamy także testu odpowiedniego do oceny ich różnic.

A teraz to samo postępowanie wobec różnic między wartościami charakteryzującymi próby A_1 i B_2 .

$$a) \bar{f}_{A_1} - \bar{f}_{B_2}$$

$$t_{\text{emp.}} = \frac{-0,109}{\sqrt{0,00865 + 0,00973}} = -0,8$$

$$b) I_{\text{oryg. } A_1} - I_{\text{oryg. } B_2}$$

Udział wyrazów o częstości 1 jest w *Dziwosłębnie* identyczny jak w *Burdzie ruskiej*, czyli różnica jest zerowa, a więc i $t_{\text{emp.}} = 0$. Prawdopodobieństwo pojawienia się wyrazu o częstości 1, 2 lub 3 natomiast wynosi w B_2 — $p_{B_2} = 0,903$. Oto sprawdzenie istotności różnicy wobec $p_{A_1} = 0,928$:

$$t_{\text{emp.}} = \frac{0,025}{\sqrt{0,000025 + 0,000371}} = 3,16$$

Interpretacja tych obliczeń jest identyczna jak przy testowaniu różnic między *Dziwosłębem* a *Żalobą*.

Wniosek wynikający bezpośrednio z dokonanych porównań — to stwierdzenie braku różnic w strukturze ilościowej słownictwa przeciwstawianych sobie tekstów. Wskaźniki świadczą o podobnym sposobie gospodarowania słowem, o leksyce bogatej, zróżnicowanej, z dużą liczbą wyrazów używanych bardzo rzadko. Pozostaje jednak teraz postawić sobie pytanie, co wynika z tego dla rozważań o autorstwie. Czy podobna struktura ilościowa słownictwa może tu świadczyć na korzyść tezy o wspólnej dla obu tekstów ręce pisarskiej? Zanim spróbujemy odpowiedzieć na to pytanie, przyjrzyjmy się jeszcze jednemu zestawieniu, zamieszczonemu na s. 184²⁸.

Tabela wskazuje niedwuznacznie, że stopień bogactwa słownikowego *Roksolanek* jest bardzo wysoki, mniej więcej tego rzędu jak w poemacie z okresu dojrzałego romantyzmu. Jeśli przy tym wspomnimy, że autor *Roksolanek* był jednym z pionierów polskiego słownictwa poezji miłosnej, że nie miał gotowego zasobu środków, z którego mógłby czerpać i przebierać w nim, to tym widoczniejsza stanie się rozrzutność

²⁷ Zob. Т. Мясник, *Математико-статистические методы в стилистике*. „Вопросы языкознания” 1967, nr 3, s. 45.

²⁸ *Słownik polszczyzny XVI wieku*. T. 1. Wrocław 1966, s. XVI–XX (dane dotyczące Reja, Kochanowskiego, Cieklińskiego i Klonowica). Dane o *Panu Tadeuszu* pochodzą z pracy Sambor *Badania statystyczne nad słownictwem*, s. 156.

Tekst	Wskaźnik <i>K</i> Kuraszkiewicza	<i>N</i>
Rej, <i>Wizerunk</i> (fragm.)	21,4	11 200
Rej, <i>Rozprawa</i>	22,0	10 710
Ciekleński, <i>Potrójny</i>	22,4	16 690
Kochanowski, <i>Pieśni</i>	26,0	14 200
Kochanowski, <i>Fraszki</i>	26,4	14 600
Klonowic, <i>Worek Judaszów</i> (fragm.)	31,4	12 900
<i>Pan Tadeusz</i> (ks. I-II)	31,13	12 172
<i>Roksolanki</i>	29,78	13 260

Zimorowica w gospodarce słowami, jego zamięrowanie do niepowtarzania wyrazu raz użytego. Skoro zaś nie udało nam się stwierdzić, mimo zastosowania czterech różnych mierników, by ta sama cecha obca była autorowi sielanek *Żaloba* i *Burda ruska*, czyli Bartłomiejowi, to nie sposób podtrzymywać tezy o rzekomo rażącej różnicy „wysłowienia”²⁹ między utworami Bartłomieja a *Roksolankami*, argumentu pełniącego w odbieraniu Bartłomiejowi autorstwa *Ruskich panien* główną rolę. Należy jednak zastrzec się, że sformułowanie „nie udało się stwierdzić różnic” nie jest równoznaczne z tym, iż różnic nie ma. Przyjęcie określonego poziomu istotności w teście weryfikującym hipotezę o braku różnic chroni bowiem przed błędem pierwszego rodzaju, czyli przed odrzuceniem hipotezy prawdziwej, natomiast bardzo skomplikowane jest postępowanie, które równie skutecznie uniemożliwiłoby popełnienie błędu drugiego rodzaju, czyli przyjęcie hipotezy fałszywej. Krótko mówiąc, pewne są tylko wnioski negatywne. Gdyby sprawdziany wykazały istotną różnicę między próbami, to wiedzielibyśmy z dużym prawdopodobieństwem ($1-a$), że ta różnica faktycznie istnieje. Skoro jednak przedstawione wyniki analiz skłaniają nas do przyjęcia hipotezy o braku różnic, to przeprowadzić musimy jak najwięcej porównań w różnych punktach, w im większej bowiem liczbie porównywanych wskaźników test nie wykaże różnic, tym mniejsze będzie ryzyko przyjęcia hipotezy fałszywej. Jest to zgodne z zasadą, że prawdopodobieństwo iloczynu zdarzeń niezależnych równe jest iloczynowi prawdopodobieństw tych zdarzeń, a więc jest tym mniejsze, im zdarzeń tych jest więcej. Stosując więc ten sposób uchronienia się przed błędem drugiego rodzaju, poddajmy teksty *A* i *B* jeszcze innemu sprawdzianowi.

Porównanie częstości występowania wyrazów gramatycznych

Wyrazy gramatyczne mają tę szczególną cechę, że obyć się bez nich nie sposób, używa się ich więc często (dzięki temu zaś już badanie próby niezbyt licznej daje wyniki wartościowe statystycznie), a przy tym częstość ich używania i ewentualna skłonność autora do nadużywania jednych, a unikania innych jest cechą stylu dość

²⁹ Zob. A. Brückner, komentarz w: Sz. Zimorowic, *Roksolanki*. Kraków 1924, s. 82. BN I 73.

głęboko ukrytą, niezauważalną (a więc jej powtórzenie, zarówno przy nieświadomym uleganiu wpływowi czyjegoś stylu, jak i świadome, przy zamierzonym naśladownictwie, jest prawie niewykonalne). Zalety badania częstości występowania najliczniejszych wyrazów gramatycznych są więc oczywiste, pod warunkiem jednak, że jest to cecha dystynktywna. Czy tak jest w istocie? Spróbujmy przyjrzeć się poniższej tabeli:

Wyraz	Roksolanki $N = 13260$		Szymonowic $N = 2545$		Różnica	Wartość sprawdzianu empirycznego
	f_R	p_R	f_S	p_S	$p_R - p_S$	$t_{\text{emp.}}$
nie	328	0,025	82	0,0322	-0,0072	-1,93
w	307	0,023	37	0,0145	0,0085	3,16
z	285	0,0215	44	0,0173	0,0042	1,47
i	240	0,018	79	0,031	-0,013	-3,59
a	108	0,0081	35	0,0138	-0,0057	-2,35
wyrazy zaprzeczone*	136	0,0102	16	0,00039	0,00981	10,96
						$ \bar{t} = 3,91 \quad s = 3,23$

* Nie są to wprawdzie w pełnym sensie wyrazy gramatyczne, ale wyrażają pewną manierę pisarską Zimorowica, warto więc poddać je badaniom.

W tabeli tej przy każdej próbie w lewej kolumnie podana jest częstość bezwzględna występowania danego wyrazu, w kolumnie prawej zaś częstość względna, czyli prawdopodobieństwo empiryczne jego wystąpienia. Dobór wyrazów do badania określony jest przez dokładność $\delta = 0,3$, którą pragniemy zachować. Ponieważ:

$$\delta = \frac{t_a}{\sqrt{N \cdot p}} = \frac{t_a}{\sqrt{f}}$$

$$f = \frac{t_a^2}{2} = \frac{(1,96)^2}{(0,3)^2} \approx 43$$

— potrzebne są takie wyrazy, których częstość wynosi przynajmniej 43. W *Roksolankach* ze względu na dużą liczebność próby prawie wszystkie wyrazy gramatyczne spełniają ten warunek, natomiast w próbie ręki Simonidesa można by wziąć pod uwagę tylko trzy wyrazy lub też zwiększyć liczebność próby. Jeśli zaś zmniejszymy żadaną dokładność tak, by nie wykraczała poza $\delta = 0,5$, to wejdą w grę również wyrazy o częstości od 15 w górę. Dodać wypada, że próba tekstu Simonidesa obejmuje sielanki *Kolacze* i *Żeńcy*, a sprawdzian empiryczny $t_{\text{emp.}}$ obliczony został podobnie jak przy porównywaniu wskaźników oryginalności i miar Gooda według wzoru:

$$t_{\text{emp.}} = \frac{p_R - p_S}{\sqrt{\frac{(1-p_R)p_R}{N_R} + \frac{(1-p_S)p_S}{N_S}}}$$

Natomiast traci częściowo wartość porównanie *Roksolanek* z próbkami ręki Simondesa i Mickiewicza, gdzie $t_{\text{emp.}}$ jest większe niż 1,96. Czyż więc częstości poszczególnych wyrazów gramatycznych są cechą odróżniającą inne ręce pisarskie? Chyba tak, skoro badanie przeprowadzone w identycznych warunkach i przy zastosowaniu tego samego, zbyt czułego testu dało wynik wskazujący na zróżnicowanie przy próbkach rzeczywiście różnych, a nie wykazało różnic przy próbkach tekstu przypuszczalnie tego samego autorstwa.

Nic więcej ponad to, że badając wszelkimi dostępnymi metodami nie wykryto różnic, powiedzieć się nie da. Uczciwsze więc będzie (jako że na rzecz autorstwa Szymona nie ma ani argumentów pewnych, ani prawdopodobnych, a autorstwo Bartłomieja wspiera w sposób bardzo prawdopodobny zarówno drobiazgowa analiza Adamczewskiego, jak i wnioskowanie statystyczne) przypisać *Roksolanki* ich właściwemu twórcy, Józefowi Bartłomiejowi Zimorowicowi.