

Jan Rybicki

## Policzmy Trylogię Sienkiewicza

## Wstęp

Czy tablice statystyczne powinniśmy traktować jak Kamienne Tablice, czy jak stoliki do seansów spirytystycznych? A czy jeżeli coś nam już powiedzą, to czy możemy im wierzyć? Jaka jest wiarygodność medium, czyli w tym wypadku statystyka, który przemawia do nas liczbami, prawie jak medium Hugo na seansach na wyspie Guernsey?<sup>1</sup>

Pytanie to jest jak najbardziej aktualne, choć Etienne Brunet postawił je już niemal dwadzieścia lat temu – co w każdej dziedzinie w najmniejszy choćby sposób powiązanej z zastosowaniami komputerowymi, jest niemal wiecznością. O ile bowiem np. językoznawcy dawno już docenili i przyzwyczaili się do korpusów, badań statystycznych i podstawowego narzędzia tych badań – komputera, o tyle zastosowania podobnych metod do badań czysto – lub choćby częściowo – literackich jest znacznie mniej akceptowane. Najczęstszy zarzut jest równocześnie bardzo prostą odpowiedzią na zadane w powyższym cytacie pytanie: czy liczby (i wykresy, na których je przedstawiamy) w ogóle coś znaczą, czy też są przypadkową gmatwaniną niepowiązanych ze sobą cyfr, wszelkie zaś graficzne ich relacje są właśnie wyłącznie takie: przypadkowe?

Ta najprostsza wypowiedź kłóci się jednak z faktem, że – również w Polsce – istnieje już od lat pięćdziesiątych skromny nurt refleksji statystyczno-literackiej w ramach literaturoznawstwa i znacznie większy w językoznawstwie; zresztą może właśnie w badaniach statystycznych na tekstem literackim najłatwiej zauważyć pojąkobsonowską jedność nauk o literaturze i nauk o języku<sup>2</sup>. Bo choć literaturoznawcy-

<sup>1</sup> E. Brunet, *What Do Statistics Tell Us?*, „Research in Humanities Computing” 1991, I, s. 70 (tłum. J.R.).

<sup>2</sup> Z zastrzeżeniem, że nie jest to ani lista kompletna, ani nawet reprezentatywna, należy wspomnieć w tym miejscu o pracach E. Engelking-Teleżyńskiej, *Z badań nad strukturą ilościową słownictwa „Vade-mecum” Cypriana Norwida*, [w:] *Język osobniczy jako przedmiot badań lingwistycznych*, red. J. Brzeziński, Zielona Góra 1988; W. Kuraszkiewicz, *Statystyczne badanie słownictwa polskich tekstów XVI wieku*, [w:] *Z polskich studiów slawistycznych*, cz. 1. *Prace językoznawcze i etnogenetyczne na IV Międzynarodowy Kongres Sławistów w Moskwie 1958*, red. P. Zwoliński, Warszawa 1958; A. Martuszewskiej, *Upraszczenie struktur (Niektóre problemy narracji powieści popularnej)*, [w:] *Studia o narracji*, red. J. Błoński, S. Jaworski, J. Sławiński, Wrocław 1982; E. Rudnickiej-Firy, *Słownictwo „Dziadów” A. Mickiewicza w świetle analizy statystycznej (wybór problematyki)*, Katowice 1986 i *Wyrazy-klucze w „Dziadach” Adama Mickie-*

-statystycy i językoznawcy-statystycy nie zawsze jeżdżą na te same konferencje, to łączą ich – co najmniej – materiały i metody: elektroniczne korpusy językowe (lub po prostu elektroniczne wersje tekstów literackich, dawniej wymagające skanowania i optycznego rozpoznawania tekstu, teraz jednak szeroko dostępne w Internecie), oprogramowanie służące do opracowania i policzenia najważniejszych wartości tekstu (w niniejszym opracowaniu posługuję się komercyjnym pakietem Wordsmith Tools 5.0) oraz oprogramowanie statystyczne (tu: Statistica 7.0).

We wcześniejszej pracy na ten temat przyrównywałem trudy i niebezpieczeństwa związane z podjęciem takich pionierskich badań do tych, które towarzyszyły Podbipięcie i Skrzetuskiemu podczas wyjścia ze Zbaraża<sup>3</sup>. Teraz zapewne docieramy już do zbawczego lasu (lasu liczb, w którym można skryć się przed wrogiem?), lecz nie znaczy to, że przedstawienie liczbowych i statystycznych cech tekstu literackiego właśnie w literackim ujęciu uzyskało już upragniony indygenat – więc chodzi tu raczej nie o działania na podobieństwo szlachetnych i szlachetnie urodzonych rycerzy Sienkiewicza, ale tego chciwego arywisty Rzędziana.

Trylogia pierwszego polskiego noblisty jest dla tak określonego zadania badawczego oczywistym i szczęśliwym wyborem. Ponieważ seria trzech książek nie tylko dostarcza statystykowi bogatego materiału, nie tylko daje okazję do porównań między poszczególnymi częściami, ale – przede wszystkim – już sam jej gatunkowy i zwyczajowy tytuł jest zaproszeniem do liczenia. Trylogia, czyli trzy książki. A skoro możemy liczyć Sienkiewiczowi książki w jego serii o siedemnastym wieku (nie zapominając przy tym, że ta liczba pojawia się również w jego „małej trylogii” i zrealizowanej tylko w jednej trzeciej trylogii z czasów Sobieskiego), to choćby w tak skromnym jak niniejsze rozpoznaniu można policzyć mu trochę więcej szczegółów.

Moim narzędziem będzie w tym wypadku policzenie i statystyczne przebadanie pewnych prostych cech badanych tekstów (przede wszystkim objętości liczonej w różnych jednostkach: słowach, zdaniach, rozdziałach) oraz pomiar bogactwa słownictwa za pomocą tzw. Standardized Type/Token Ratio (STTR, uśredniony iloraz *type* i *token*). Wielkości te będą dodatkowo powiązane z innymi rozpoznaniem, przede wszystkim z wyszukaniem statystycznych słów kluczowych. W opinii wielu badaczy bowiem najciekawsze wyniki otrzymujemy wtedy, gdy łączymy kilka metod w jednej. Jak twierdzi w swej najnowszej, niepublikowanej jeszcze pracy jeden z guru amerykańskiej stylometrii komputerowej, David Hoover:

Policzyć słowa, policzyć TTR, zbadać rozkład tej czy innej cechy, przetworzyć dane przez akrusz kalkulacyjny Deltę – to samo w sobie jest dość ciekawe. Ale naprawdę mia-

---

wicza odzwierciedleniem postaw ideowo-artystycznych poety, „Rocznik Naukowo-Dydaktyczny WSP w Krakowie, z. 168: Prace Językoznawcze VIII”, Kraków 1994; J. Sambor, *Badania statystyczne nad słownictwem (na materiale „Pana Tadeusza”)*, Wrocław 1969 i *Analiza stosunku „Type-Token”, czyli objętości słownictwa (W) do długości tekstu (N). Na przykładzie tekstu „Pana Tadeusza”*, „Prace Filologiczne” 1970, t. 20; T. Smółkowej, *Słownictwo i fleksja „Lalki” Bolesława Prusa. Badania statystyczne*, Wrocław 1974; E. Stachurskiego, *Słownictwo w utworach polskich naturalistów. Badania statystyczne*, Kraków 1989, *Rozwój i zróżnicowanie słownictwa pisarza. Na podstawie prób tekstowych z utworów Adolfa Dygasińskiego*, Kraków 1993 i *Słowa-klucze polskiej epiki romantycznej*, Kraków 1998; E. Zembaty-Michalakowej, *Poezja Juliana Przybosa w świetle badań statystyczno-językowych na tle porównawczym*, Wrocław 1982.

<sup>3</sup> J. Rybicki, „A Computer Assisted Comparative Analysis of Two English Translations of Henryk Sienkiewicz's Trilogia”, rozprawa doktorska, Akademia Pedagogiczna w Krakowie, Kraków 2000, s. 186.

rodajne i atrakcyjne dla tradycyjnego literaturoznawcy wyniki otrzymamy wtedy, gdy nasze i jego wnioski opierać się będą łącznie na wszystkim, co da się policzyć i powiązać z tekstem i jego interpretacyjnym odczytaniem. Obliczenie częstości najczęściej występujących słów nie pozwala do końca podzielić twórczości Jamesa na etapy stylistyczne; dopiero nałożenie kilku lub kilkunastu parametrów daje to, co chcemy uzyskać<sup>4</sup>.

## Księgi Trylogii

Zacznijmy od wartości rzeczywiście najprostszych: od rozmiarów arcydzieła Sienkiewicza. Wiemy oczywiście, że Trylogia jest długa – wpływowy i złośliwy angielski krytyk Edmund Gosse użył określenia „słoniowaty romans”<sup>5</sup> – ale dokładnie jak długa? Zatem na przykład: ile różnych słów (*word-types*) używa Sienkiewicz (inaczej: jakich rozmiarów jest słownik użytych przezeń wyrazów), z jaką częstością, i ile jest tych słów wszystkich razem (*word-tokens*)? Albo: ile jest w Trylogii i w jej różnych częściach narracji, a ile dialogu? I czy poszczególne jej części zachowują podobne proporcje?

**Tab. 1.** Dane objętości narracji i dialogu w trzech księgach Trylogii

Tytuł	Forma podawcza	Liczba słów	Narracja/Dialog
<i>Ogniem i Mieczem</i>	Narracja	145 065	60,69%
	Dialog	93 975	39,31%
	łącznie	239 040	–
<i>Potop</i>	Narracja	208 238	54,72%
	Dialog	172 286	45,28%
	łącznie	380 524	–
<i>Pan Wołodyjowski</i>	Narracja	97 004	60,72%
	Dialog	62 745	39,28%
	łącznie	159 749	–
<i>Trylogia</i>	Narracja	450 307	57,81%
	Dialog	329 006	42,19%
	łącznie	779 313	–

Tabela 1 pokazuje więc, że Trylogia Henryka Sienkiewicza ma objętość *dokładnie* 779 313 słów, z czego 239 040 to długość pierwszej części, 380 524 – drugiej i 159 749 – trzeciej. Rzecz jasna nie chodzi tu o „odkrycie” (które za pomocą choćby popularnego edytora Microsoft Word może dokonać na nowo właściciel każdej z dostępnych w Internecie elektronicznych wersji powieści), że *Potop* jest dłuższy od *Ogniem i mieczem*, ta zaś część Trylogii od *Pana Wołodyjowskiego*. Chodzi o konkretne opisanie rzeczywistości. Równie ciekawy jest prosty fakt, że Sienkiewicz używa w kolejnych częściach Trylogii 68 309 różnych słów i form gramatycznych (*types*).

Z tabeli 1 widać również, że poszczególne księgi Trylogii nie zachowują dokładnie tych samych proporcji między dialogiem a narracją: postaci Sienkiewicza są najbardziej gadatliwe w *Potopie*, z kolei najwięcej narracji – a więc opisów – znajdujemy

<sup>4</sup> D. Hoover, „Modes of Composition in Henry James: Dictation, Style, and What Maisie Knew”, referat na konferencji Digital Humanities 2009, College Park 2009 (tłum. J.R.).

<sup>5</sup> E.E. Gosse, *Henryk Sienkiewicz*, „Contemporary Review” 1897, LXXI, s. 513–525.

w *Ogniem i mieczem*. O ile jednak humaniście wystarczy, że proporcje dialogu do narracji tak bardzo różnią się między *Potopem* a pozostałymi częściami Trylogii, o tyle statystyk musi obliczyć, czy dla takich danych różnica np. między ową proporcją jest statystycznie znacząca – czyli na tyle duża, że nie może być dziełem przypadku.

Klasyczną metodą obliczania owej istotności statystycznej jest tzw. test z, obliczany według następującego wzoru:

$$z = \frac{i_1 - i_2}{s_{i_1 - i_2}}$$

gdzie  $i_1$  to iloraz liczby *tokens* narracji do *tokens* dialogu w jednej z książek,  $i_2$  – w drugiej;  $s_{i_1 - i_2}$  to błąd standardowy różnicy między oboma ilorazami, obliczany z kolei według formuły:

$$s_{i_1 - i_2} = \sqrt{\frac{i(1-i)}{n_1} + \frac{i(1-i)}{n_2}}$$

gdzie  $i$  to średnia ważona obu ilorazów, zaś  $n_1$  i  $n_2$  to całkowita objętość (również w *tokens*) odpowiednio pierwszej i drugiej książki.

Tak uzyskany wynik testu z jest następnie porównywany z tabelą prawdopodobieństwa (tu: prawdopodobieństwa przypadkowości uzyskania takich różnic między proporcjami narracji i dialogu). Możemy mówić o nieprzypadkowości, a więc o istotności statystycznej badanych różnic, gdy prawdopodobieństwo to jest mniejsze od 0,05.

Dopiero na takiej podstawie możemy orzekać, że większa gadatliwość bohaterów *Potopu* jest istotnie (w znaczeniu statystycznym) zjawiskiem potwierdzonym, a także że ta sama cecha postaci w skrajnych częściach sienkiewiczowskiego cyklu statystycznie istotna nie jest. Dokładne wyniki przedstawia tabela 2.

**Tab. 2.** Proporcje narracji i dialogu w trzech księgach Trylogii wraz z istotnością statystyczną różnic

Tytuł	Proporcje	Test z	p	Istotność statystyczna
<i>Ogniem i mieczem/Potop</i>	60,69%	46,15	0	TAK
<i>Potop/Pan Wołodyjowski</i>	54,72%	-40,59	0	TAK
<i>Pan Wołodyjowski/Ogniem i mieczem</i>	60,72%	0,23	0,89	NIE

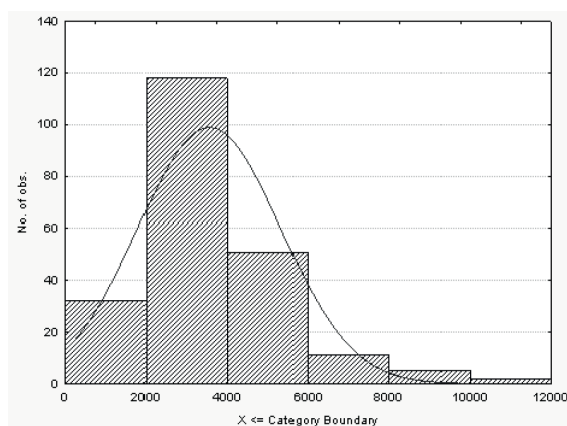
## Długość rozdziałów

Ale Trylogia to nie tylko trzy książki, ale również 219 rozdziałów. Właśnie rozdziały stanowią kolejny poziom, który może mieć znaczenie przy interpretacji dzieła. W przypadku cyklu Sienkiewicza jest to o tyle ważne, że rozdziały, tworząc podstawową, liniową strukturę narracyjną dzieła, najczęściej przedstawiają wydarzenia powiązane ze sobą wspólnym miejscem akcji, wątkiem czy etapem rozwoju narracji. Długość rozdziału może więc być wyznacznikiem sensu (czy choćby charakteru) danego elementu narracyjnego. Równocześnie jednak długość rozdziałów w danej części Trylogii lub w jej całym tekście może być uważana za element stylu, badanie zmian tej wartości w przebiegu cyklu to interesujący wgląd w ewolucję tego aspektu pisarstwa Sienkiewicza.

**Tab. 3.** Objętość rozdziałów w Trylogii (mierzona liczbą słów)

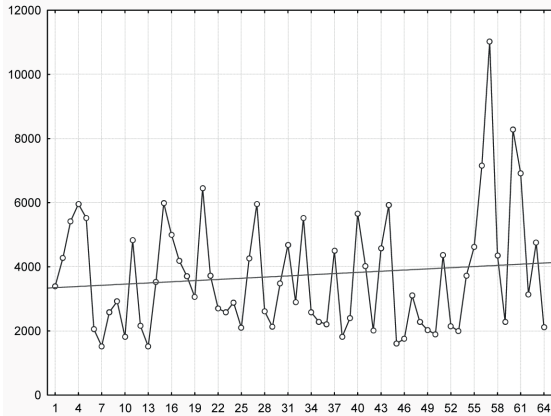
Liczba rozdziałów	219
Średnia objętość	3558,5
Najdłuższy rozdział (57)	11021
Najkrótszy rozdział (162)	458
Standardowe odchylenie	1768,1

Tekst najdłuższego z wszystkich rozdziałów Trylogii (57 dla całej Trylogii) to opis początku oblężenia Zbaraza – od przybycia Zagłoby i Wołodyjowskiego do pierwszego szturm. Najkrótszym jest wstęp do *Pana Wołodyjowskiego* (167), stanowiący li tylko przypomnienie sytuacji, w której czytelnik – stosunkowo niedawno – rozstał się z bohaterami *Potopu*. Rozkład objętości rozdziałów w całej Trylogii przedstawia się następująco (wykres 1):

**Wykres 1.** Rozkład objętości rozdziałów w Trylogii

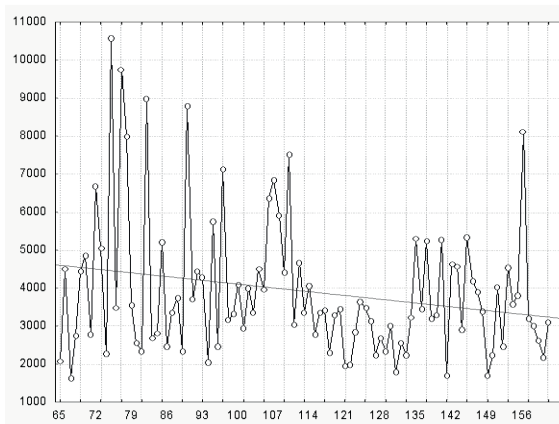
Jak widać, najwięcej rozdziałów (118) znajduje się przedziale 2000–4000 słów, a więc dokładnie wokół średniej. Drugim „popularnym” przedziałem jest następny: 4000–6000 słów (51). Łącznie w przedziałach tych mieści się więc 169 rozdziałów, czyli prawie 77%. Z kolei rozdziały bardzo długie (powyżej 8000 słów) zdarzają się w całej Trylogii tylko siedmiokrotnie (3,2%). Są to (oprócz wspomnianego już „rekordzisty”) rozdziały 75 (10 584 słów: opis początku potopu szwedzkiego w Wielkopolsce włącznie z kapitulacją pod Ujściem), 77 (9762: początek pobytu Zagłoby i Skrzetuskich na dworze Janusza Radziwiłła w Kiejdanach), 82 (8979: uwolnienie pułkowników przez Zagłobę), 90 (8795: podróż Kmicica przez Podlasie aż do porwania księcia Bogusława), 60 (8278: wyjście ze Zbaraza i śmierć Podbiپیęty) oraz 156 (8138: bitwa pod Prostkami). Bliski umownej granicy 8000 słów jest jeszcze jeden rozdział: 78 (7999: ucztą kiejdańska). Jak widać, są to w większości obszernie sceny batalistyczne (75, 90, 60, 156) lub przynajmniej zawierające elementy walki. Charakterystycznym, choć raczej spodziewanym zjawiskiem jest przewaga wśród najdłuższych rozdziałów fragmentów *Potopu* oraz ich brak w *Panu Wołodyjowskim*. Co ciekawe, wśród tych najdłuższych rozdziałów nie znalazł się żaden z kończących dwie spośród trzech części Trylogii „wielkich” opisów batalistycznych.

Powyższe zestawienie zdaje się sugerować, że długie rozdziały zdarzają się w Trylogii częściej w początkowych i końcowych partiach każdej części. Aby dokładniej przeanalizować tę zależność, należy sporządzić wykresy długości rozdziałów w zależności od miejsca w tekście.



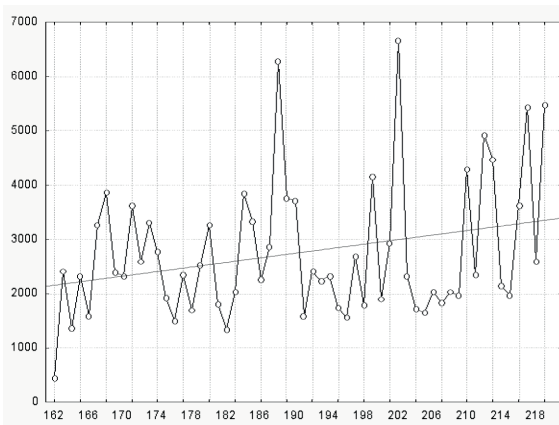
Wykres 2. Długość rozdziałów w *Ogniem i mieczem*

Wykres 2 ukazuje ciekawe zjawisko w *Ogniem i mieczem*: tendencje wzrostowe w długości rozdziałów realizowane są znacznie szybciej niż spadkowe. Inaczej mówiąc, o ile lokalne maksimum wykresu osiągnięte jest w jednym, najwyżej dwóch etapach, spadek często rozłożony jest na kilka kolejnych rozdziałów. Znaczącym wyjątkiem od tej reguły jest stopniowe narastanie długości rozdziałów prowadzące do maksimum pod koniec powieści, w którym – zgodnie ze zgłoszoną powyżej hipotezą – grupują się najdłuższe rozdziały *Ogniem i mieczem*. Zjawisko to jest też ukazane przez lekko wznoszący się ogólny trend. W przedziale 2000–6000 słów mieści się w tej części Trylogii 59 rozdziałów (ponad 92%), co wskazuje na bardzo równomierne wartości w tej części w porównaniu z całością cyklu.



Wykres 3. Długość rozdziałów w *Potopie*

Zupełnie inaczej przedstawia się struktura długości rozdziałów w *Potopie* (wykres 3). Niemal do połowy książki występują bardzo silne amplitudy: sąsiednie rozdziały różnią się często między sobą o ponad 6000 słów, tam też znajduje się najwięcej rozdziałów powyżej 8000 słów długości. „Uspokojenie” przynosi trzeci kwartał powieści: rozdziały stają się krótkie, amplitudy nieznaczne, w ostatniej ćwierci tekstu amplitudy znów rosną (choć nie osiągają wartości z pierwszej połowy powieści). Nic zaskakuje, że ogólny trend jest wyraźnie opadający. W tej sytuacji nie zaskakuje też to, że w przedziale 2000–6000 znajduje się 82% rozdziałów – a więc mniej niż w *Ogniem i mieczem*.

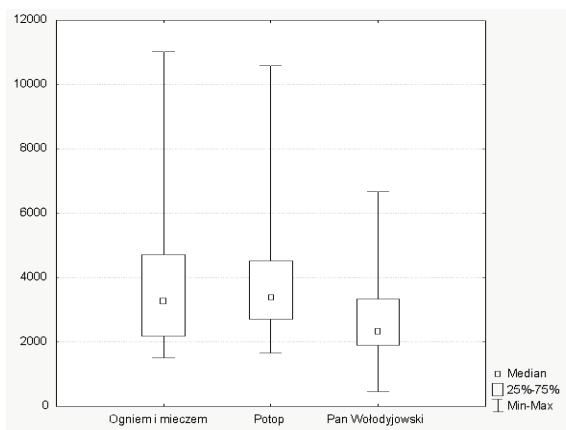


**Wykres 4.** Długości rozdziałów w *Panu Wołodyjowskim*

Jeszcze inny obraz przynosi wykres 4, ukazujący długości rozdziałów w ostatniej części Trylogii. Trend jest tu wyraźnie wzrostowy, ale dwie najwyższe wartości – choć nieprzekraczające przyjętego wcześniej progu 8000 słów – znajdują się tym razem w połowie utworu. Chodzi o rozdziały 188 (6280: bitwa z Azba-bejem) i 202 (6676: ucieczka Basi do Chreptiowa) – a więc znów o partie tekstu obfitujące w sceny walki i dramatyczne wydarzenia. W tej perspektywie wydaje się oczywiste, że kolejna fala lokalnych wzrostów pojawi się wraz z początkiem oblężenia Kamieńca i zakończy trzecim co do długości rozdziałem opisującym bitwę chocimską.

Ogólny obraz długości rozdziałów w Trylogii – tym razem bez odniesień do miejsca w całości powieści – można ukazać w znacznie prostszy sposób za pomocą tzw. wykresu pudełkowego. Wykres 5 ukazuje te wahania dla wszystkich trzech części cyklu: mediana to wartość, powyżej i poniżej której znajduje się 50% pomiarów (tu: długości rozdziałów), „pudełko” określa zakres połowy wartości wokół mediany, czyli tzw. rozstęp ćwiartkowy, dolny i górny koniec odcinka to wartości skrajne. Ten wykres pozwala uzupełnić obserwacje na temat rozkładu objętości rozdziałów Trylogii o następujące spostrzeżenia: ogólnie największa różnorodność tej wartości występuje w pierwszej, najmniejsza w trzeciej części cyklu, *Pan Wołodyjowski* wykazuje również wyraźne przesunięcie w dół w stosunku do pozostałych powieści. Wykres ten pozwala ustalić, że trzecia część Trylogii istotnie różni się pod tym względem.

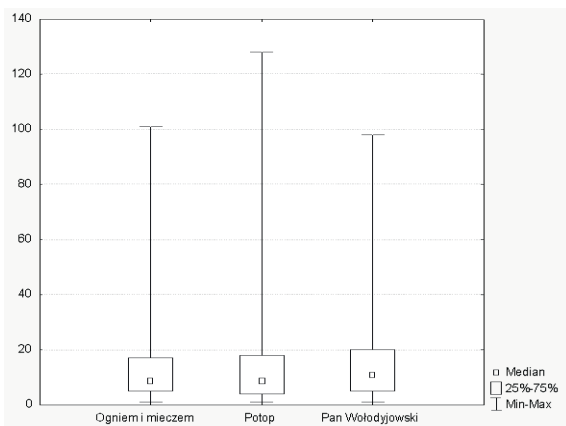




**Wykres 5.** Długości rozdziałów w trzech częściach Trylogii (mediana, rozstęp ćwiartkowy, wartości skrajne)

## Długości zdań

Od analogicznego wykresu warto rozpocząć rozważania na temat długości zdań w sienkiewiczowskim cyklu (wykres 6):



**Wykres 6.** Długość zdań w Trylogii (mediana, rozstęp ćwiartkowy, wartości skrajne)

Jak można się było spodziewać, średnia długość zdań nie różni się wiele we wszystkich trzech częściach Trylogii. Jedyną znaczącą różnicą jest tutaj długość najdłuższego zdania: o ile w skrajnych powieściach jest to wartość na poziomie 100 słów, o tyle w *Potopie* najdłuższe zdanie ma ich aż 129. Jednak, jak wynika z wykresu oraz danych przedstawionych w tabeli 4, we wszystkich trzech częściach Trylogii zdecydowana większość zdań jest nie dłuższa niż 20 słów, w 40 słowach mieści się ich ponad 95%. Warto jednak zauważyć, że *Pan Wołodyjowski* odbiega nieco od reszty cyklu pod względem proporcji między pierwszym a drugim przedziałem długości zdań.



Tab. 4. Przedziały długości zdań w trzech częściach Trylogii

Przedział (liczba słów)	<i>Ogniem i mieczem</i>		<i>Potop</i>		<i>Pan Wołodyjowski</i>	
	Liczba zdań	Procent całości	Liczba zdań	Procent całości	Liczba zdań	Procent całości
0–20	15953	80,78286	25597	80,99035	8902	75,84562
20–40	3253	16,47255	5210	16,48473	2414	20,56744
40–60	490	2,48126	701	2,21800	370	3,15242
60–80	44	0,22281	87	0,27527	48	0,40896
80–100	7	0,03545	7	0,02215	3	0,02556
100–120	1	0,00506	2	0,00633	–	–
120–140	–	–	1	0,00316	–	–

Przy okazji warto zidentyfikować te najdłuższe zdania w każdej z części powieści. Najdłuższe zdanie *Ogniem i mieczem* opisuje oczarowanie pana Wołodyjowskiego Heleną podczas powrotu znad Waładynki:

Wobec tamtych [kobiet – J.R.] był pan Wołodyjowski i ochoczy, i mowny, a teraz, gdy spoglądał na te oczy aksamitne, słodkie a mdlejące, na te jedwabne ich zasłony, których cień padał aż na jagody, na ten włos rozsypany, jakby kwiat hiacyntowy, po ramionach i plecach, na strzelistość postaci, na pierś wypukłą i tchnieniem lekko kołysaną, od której biło ciepło lubę, na te wszystkie białości liliowe i róże a maliny ust – gdy na to wszystko spoglądał pan Wołodyjowski, wówczas po prostu języka w gębie zapominał i, co najgorsza, że się sobie wydawał niezgrabny, głupi, a zwłaszcza, mały, ale to tak mały, że aż śmieszny. (101 słów)

Jest ono jednak wyjątkiem wśród najdłuższych zdań tej powieści, inne długie (i niewiele od najdłuższego krótsze) zdania dotyczą już nie uczuć bohaterów ani nawet nie scen batalistycznych (jak w przypadku najdłuższych rozdziałów), ale przede wszystkim polityki, i to w odniesieniu do trzech głównych postaci tego właśnie wątku *Ogniem i mieczem*:

Poselstwa zagraniczne były milczącym uznaniem zarówno jego potęgi, jak udzielności; stała przyjaźń Tatarów, opłacana większością zdobytych łupów i nieszczęsnym jasyrem; który ten wódz ludowy z ludu wybierać pozwolił – obiecywała poparcie przeciw każdemu nieprzyjacielowi; dlatego to Chmielnicki, uznający jeszcze pod Zamościem zwierzchnictwo i wolę królewską, obecnie wbity w pychę, przekonany o swej sile, o nieładzie Rzeczypospolitej, niedołęstwie jej wodzów, gotów był i na samego króla podnieść rękę marząc już teraz w posępnej swej duszy nie o swobodach kozackich, nie o powrocie dawnych przywilejów Zaporozu, nie o sprawiedliwości dla siebie, lecz o państwie udzielnym, o czapce książęcej i berle. (98)

On to przecie, jako Rusin z krwi i kości pierwszy wziął na się rolę pacyfikatora w tej bezprzykładnej wojnie – on występował wszędy, w senacie i na sejmie, jako najgorętszy stronnik układów, on popierał politykę kanclerza i prymasa, on potępiał najsilniej Jeremiego i działał w dobrej wierze dla dobra kozactwa i Rzeczypospolitej – i wierzył całą swą gorącą duszą, że układy, że ustępstwa wszystko pogodzą, uspokoją, zbliznią – i właśnie teraz, w tej chwili, gdy wiózł buławę Chmielnickiemu, a ustępstwa Kozaczyż-

nie, zwątpił o wszystkim – ujrzał oczywiście marność swoich wysileń, ujrzał pod nogami próżnię i przepaść. (96)

I nie tej fortuny olbrzymiej, równej całym księstwom niemieckim, żałował książę, ale się do tego dzieła rąk własnych przywiązał; wiedział, że gdy jego tu zbraknie, wszystkiego zbraknie, że praca lat całych od razu zostanie zniszczona, że trud pójdzie na marne, dzicz się rozpęta, pożary ogarną wsie i miasta, że Tatar będzie poił konie w tych rzekach, bór porośnie na zgliszczach i że jeśli Bóg da wrócić – wszystko, wszystko wypadnie poczynać na nowo – a może już tych sił nie będzie i czasu zbraknie, i ufności takiej, jak pierwiej, nie stanie. (91)

Brak wśród najdłuższych zdań scen walki wynagradza *Potop*, którego najdłuższe zdanie ukazuje liczne zagraniczne wyczyny księcia Bogusława podczas wstępnego opisu czarnego charakteru obu wątków *Potopu*, miłosnego i publicznego:

On to w Holandii rzucał się w największy to war bitwy, między niezrównane pułki piechoty hiszpańskiej, i własną książęcą ręką zdobywał chorągwie i działa; on na czele regimentów księcia Oranii zdobywał baterie, przez starych wodzów uznane za niepodobne do zdobycia; on nad Renem, na czele muszkietierów francuskich, rozbił ciężkie chorągwie niemieckie, w trzydziestoletniej wojnie wyćwiczone: on ranił w pojedynku, we Francji, najsłynniejszego między kawalerami francuskimi fehmistrza, księcia de Fremouille; drugi słynny zabijaka, baron von Goetz, prosił go na kłęczkach o darowanie życia; on ranił barona Grota, za co musiał od brata Janusza słuchać gorzkich wymówek, iż pospolituje godność swą książęcą stawając do walki z nierównego stanu ludźmi; on wreszcie, wobec całego dworu francuskiego, na balu w Luwrze uderzył w twarz margra-biego de Rieux za to, iż mu „szpetnie” przymówił. (129)

Również drugie co do długości zdanie jest poświęcone bojowej sprawności jednej z głównych postaci, tym razem w bezpośrednim już opisie batalistycznym. Tak oto poczyna sobie pan Wołodyjowski podczas swej pierwszej bitwy ze Szwedami w Klewanach:

I co chwila jakiś kapeluszczyk szwedzki zapadał przed nim w ciżbę, jakoby nurka dawał pod ziemię; czasami rapier, wytrącony z rąk rajtara, wylatywał furkoczając nad szereg, a jednocześnie odzywał się krzyk ludzki przeraźliwy i znów kapeluszczyk zapadał; zastępował go drugi, drugiego trzeci, lecz pan Wołodyjowski posuwał się ciągle naprzód, małe jego oczki świeciły jak dwie skry złowrogie, i nie unosił się, i nie zapamiętywał, nie machał szablą jak cepem; chwilami, gdy nie miał nikogo na długość szabli przed sobą, zwracał twarz i klingę nieco w prawo lub w lewo i strącał w mgnieniu oka rajtara ruchem na pozór nieznacznym, i straszny był przez te ruchy małe, a błyskawiczne, prawie niezłowiecze. (110)

Co ciekawe, dwa inne z najdłuższych zdań *Potopu* też poruszają kwestie militarne, jednak na poziomie bardziej strategicznym. Opis niewesołej sytuacji wyprawy Karola Gustawa w głąb Polski jest już klasycznym homeryckim porównaniem:

Ale jak gdy wśród gęszczy borów potężny niedźwiedź niesie swe ciężkie cielsko, krusząc po drodze krze i gałęzie, wilcy zaś idą w trop za nim i nie śmiąc mu drogi zastąpić, coraz

bliżej następują nań z tyłu, tak i owe partie ciągnęły za armią Karola, w coraz ciasniejsze łącząc się gromady, i szły za Szwedem, jak cień idzie za człowiekiem, i wytrwale jak cień, bo we dnie i w nocy, w pogodę i w niepogodę; przed nim zaś psuto mosty, niszczone zapasy, że musiał iść jak w pustynię, nie mając głowy gdzie schronić lub się w głódzie czym pokrzepić. (99)

Podobny i stylistycznie, i treściowo opis powtarza Sienkiewicz nieco później, gdy przedstawia wiadomości, jakie dobiegają z kraju do Taurogów – też zresztą w kontekście wyprawy króla Szwecji:

Ile było drzew w borach Rzeczypospolitej, ile kłosów kołysało się na jej łąkach, ile gwiazd świeciło po nocach między Tatarami a Bałtykiem, tyle wstało przeciw Szwedom wojowników: którzy szlachtą będąc, do miecza a wojny z woli bożej i przyrodzonego rzeczy porządku się rodzili; którzy skiby pługiem krając, obsiewali ziarnem tę krainę; którzy handlem i rzemiosły po miastach się parali; którzy żyli w puszczech z pszczelnej pracy, z wypalania smoły, z topora lub strzelby; którzy nad rzekami siedząc, rybactwem się trudnili; którzy na stepach koczowali ze stadami – wszyscy chwycili za broń, aby najezdnika z kraju wyżęnać. (96)

Bardzo charakterystyczne są trzy najdłuższe zdania w *Panu Wołodyjowskim*, wszystkie bowiem dotyczą kontekstu kresowego. Dwa z nich stanowią zresztą opis mniej lub bardziej realnych planów ucywilizowania tej polskiej *frontier*. W najdłuższym zdaniu tej części opisana jest „pozytywistyczna” działalność Wołodyjowskiego w okolicach Chreptiowa:

Obecnie były to dopiero początki, siła jeszcze pozostawało do roboty; drogi nie były jeszcze bezpieczne; rozwydrzony lud chętniej ze zbójcami niż z wojskiem w komitywę wchodził i za lada przyczyną znów krył się w gardziele skaliste; przez Dniestrowe brody często przekradały się watahy złożone z Wołochów, Kozaków, Węgrzynów, Tatarów i Bóg wie nie kogo; te zapuszczały zagony po kraju, napadając po tatarsku wsie, miasteczka i zgarniając wszystko, co się zgarnąć dało; chwili jeszcze nie można było w tych stronach szabli z ręki popuścić ani muszkietu na gwoździu zawiesić, jednakże początek już był uczyniony i przyszłość zapowiadała się pomyślnie. (98)

Bardziej dalekosiężną, ale i bardziej utopijną wizję przyszłości Kresów przedstawia w imieniu Azji zarażony nią Bogusz podczas rozmowy z Sobieskim:

Przed zdumionymi oczyma wielkiego hetmana przesuwają się jakoby jasne obrazy przyszłości: więc tysiące i dziesiątki tysięcy Tatarów ciągnęły wraz z żonami, z dziećmi i ze stadami na ziemię i wole; więc przerażeni Kozacy, widząc tę nową siłę Rzeczypospolitej, bili kornie czołem przed nią, przed królem i przed hetmanem; więc nie było już więcej buntów na Ukrainie, więc starymi szlakami nie szły niszczące jak płomień lub powódź zagony na Ruś, a natomiast obok wojsk polskich i kozackich buszowały po niezmiernych stepach, z graniem trąb i hukiem kotłów, czambuły ukraińskiej szlachty – Tatarów... (91)

Z kolei przeszłość wschodnich granic Rzeczypospolitej jest tematem wstępu do gawęd „rycerzy spod stepowych stanic” w chreptiowskiej świetlicy:

A mieli co opowiadać, byli bowiem między nimi tacy, którzy odwiedzili Szwecję i Moskwę; byli tacy, którzy młode lata na Siczy jeszcze przed chmielniczczyzną spędzili; byli, którzy swego czasu w Krymie owiec jako niewolnicy strzegli; którzy w Bachczysaraju studnie, w niewoli będąc, kopali; którzy zwiedzili Azję Mniejszą; którzy po Archipelagu na galerach tureckich wiosłowali; którzy w Jerozolimie czołem o grób Chrystusa bili; którzy doświadczyli wszelkich przygód i wszelkich niedoli, a przecie jeszcze pod chorągwie wrócili bronić do końca życia, do ostatniego tchu tych krain pogrzeźnych, krwią zlanych. (87)

W tym miejscu bardzo interesująca mogłaby być analiza sekwencyjna układów zdań i ich długości w tekście *Potopu* – choćby przypominająca tę z poprzedniej części niniejszego opracowania – ale poziom złożoności tego zagadnienia jest znacznie wyższy niż w przypadku rozdziałów. Prosta statystyka opisowa raczej nie może tu mieć zastosowania, bardzo obiecujące wyniki dają natomiast metody wykorzystujące różne pojęcie szeregów czasowych dla badania „rytmu” zmian danej wartości oparte na różnych modelach autoregresyjnych, a stosowane również w badaniach polskich autorów – nie tylko w zdaniu, lecz może przede wszystkim w najróżniejszych systemach metrycznych wiersza<sup>6</sup>.

### Słowa w dialogach

Nie trzeba wcale używać komputera – a nawet nie trzeba umieć liczyć – by wiedzieć, że najwięcej w całej Trylogii mówi Zagłoba. W tabeli 5 umieszczono 47 postaci Trylogii, których partie dialogowe przekraczają poziom 1000 słów, ukazuje ona, do jakiego stopnia jest to prawda: Zagłoba wypowiada niemal co piąte słowo dialogu w trzech powieściach Sienkiewicza. I tu nikt nie może się z nim równać, bo następni bohaterowie w tej klasyfikacji, Kmicic i Wołodyjowski, prawie że dzielą się drugim miejscem, ale już z wynikiem o połowę gorszym od prowadzącego. W czołówce dominują poza tym postacie, które łączy – i nie ma w tym nic dziwnego – przynależność do środowiska „męsko-polsko-szlacheckiego”. Równie mało zaskakujący jest fakt, iż pierwsza kobieta na tej liście zajmuje dopiero miejsce 9 (Basia). Co prawda Oleńka jest niedaleko, bo na miejscu 12, ale trzecią z głównych bohaterek, Helenę, znaleźć można dopiero na miejscu 39 – co należy porównać z wysokimi miejscami, jakie zajmują w tym rankingu ich partnerzy w głównych wątkach miłosnych cyklu.

Jeszcze większa niż w całej Trylogii jest przewaga Zagłoby w *Ogniem i mieczem*, w której to powieści wypowiada on już niemal co trzecie słowo dialogu. To równocześnie niemal trzy razy więcej niż rozmiary partii Skrzetuskiego i pięć razy więcej niż mówi Rzędzian, Wołodyjowski lub Chmielnicki. Pierwszy z tej trójki zawdzięcza tę wysoką trzecią pozycję swej roli informatora o części wydarzeń powieści wraz z licznymi i zabawnymi dygresjami; dla Wołodyjowskiego jest to najniższy wynik w całej Trylogii, ale też postać ta dopiero wyłania się z tłumu wielu innych oficerów Jaremy, z kolei Chmielnicki jest głównym wyrazicielem i reprezentantem postaw

<sup>6</sup> Przede wszystkim: A. Pawłowski, *Metody kwantytatywne w sekwencyjnej analizie tekstu*, Warszawa 2001, i wcześniejsze artykuły tegoż autora, ale również M. Eder, *How Rhythmic is Hexameter: A Statistical Approach to Ancient Epic Poetry*, *Proc. Digital Humanities 2008*, Oulu 2008.

ukraińskich. Jeżeli do owych pięciu bohaterów dodamy jeszcze sąsiadujące w rankingu partie Bohuna i Jeremiego Wiśniowieckiego, okaże się, iż na pozostałe 19 postaci występujących w *Ogniem i mieczem* pozostaje już tylko łącznie nieco ponad 32% dialogu – z czego trzeba obdzielić nie tylko pomniejsze postaci męskie, nie tylko małomówną, bo bardzo pasywną postać Heleny, ale i bardziej od niej gadatliwą, choć pojawiającą się tylko na początku powieści Kurcewiczową, oraz bardziej od kniaziówny epizodyczną, ale i bardzo krwistą kreację Horpyny. A choć o znaczeniu danej postaci dla całości powieści nie decyduje wyłącznie liczba wypowiedzianych przez nią słów, to jednak przykład przyszłej pani Skrzetuskiej dość dobrze ilustruje rolę kobiet w „męskich” powieściach przygodowych Sienkiewicza<sup>7</sup>.

**Tab. 5.** Rozmiary partii poszczególnych postaci Trylogii

Lp.	Postać	<i>Ogniem i mieczem</i>		<i>Potop</i>		<i>Pan Wołodyjowski</i>		Trylogia	
1	Zagłoba	27 791	29,57%	19 731	11,45%	14960	23,84%	62482,41	18,99%
2	Kmicic	–	–	32 249	18,72%	526	0,84%	32775,19	9,96%
3	Wołodyjowski	6229	6,63%	14 548	8,44%	11325	18,05%	32102,15	9,76%
4	Skrzetuski	10 169	10,82%	2048	1,19%	29	0,05%	12246,12	3,72%
5	Bogusław	–	–	10421	6,05%	124	0,20%	10545,06	3,21%
6	Janusz Radziwiłł	–	–	10 475	6,08%	–	–	10475,06	3,18%
7	Rzędzian	6350	6,76%	2106	1,22%	–	–	8456,08	2,57%
8	Jan Kazimierz	376	0,40%	5591	3,25%	–	–	5967,036	1,81%
9	Basia	–	–	–	–	5446	8,68%	5446	1,66%
10	Ketling	–	–	2761	1,60%	2250	3,59%	5011,016	1,52%
11	Chmielnicki	4958	5,28%	–	–	–	–	4958,053	1,51%
12	Oleńka	–	–	4637	2,69%	251	0,40%	4888,027	1,49%
13	Charlamp	397	0,42%	3536	2,05%	940	1,50%	4873,025	1,48%
14	Bohun	4236	4,51%	–	–	–	–	4236,045	1,29%
15	Jeremi Wiśniowiecki	4193	4,46%	–	–	–	–	4193,045	1,27%
16	Azja	–	–	–	–	3881	6,19%	3881	1,18%
17	Miecznik	–	–	3633	2,11%	–	–	3633,021	1,10%
18	Longinus	3345	3,56%	–	–	–	–	3345,036	1,02%
19	Muszalski	–	–	–	–	3010	4,80%	3010	0,91%
20	Kiemlicz	–	–	2711	1,57%	–	–	2711,016	0,82%
21	Kordecki	–	–	2707	1,57%	–	–	2707,016	0,82%
22	Sapieha	–	–	2687	1,56%	–	–	2687,016	0,82%
23	Anusia	434	0,46%	2156	1,25%	–	–	2590,017	0,79%
24	Sobieski	–	–	–	–	2575	4,10%	2575	0,78%

<sup>7</sup> Nie inaczej jest w *Quo vadis*, gdzie Ligia, co najmniej równie silny motor działań głównego bohatera, ma do dyspozycji niespełna 2% dialogu, mniej nawet od Akte (K. Czubak, „A Computer-assisted Comparative Stylistic-Statistical Analysis of English and Hungarian Translations of Henryk Sienkiewicz’s *Quo vadis*”, praca magisterska, Akademia Pedagogiczna w Krakowie, Kraków 2008, s. 90). Niczego innego nie należy się spodziewać po partiach Nel czy nawet Maryni.

Lp.	Postać	<i>Ogniem i mieczem</i>		<i>Potop</i>		<i>Pan Wołodyjowski</i>		Trylogia	
25	Zaćwilichowski	2567	2,73%	–	–	–	–	2567,027	0,78%
26	Sakowicz	–	–	2388	1,39%	–	–	2388,014	0,73%
27	Bogusz	–	–	–	–	1937	3,09%	1937	0,59%
28	Nowowiejski	–	–	–	–	1882	3,00%	1882	0,57%
29	Zamoyski	–	–	1872	1,09%	–	–	1872,011	0,57%
30	Kamiński	–	–	–	0,00%	1844	2,94%	1844	0,56%
31	St. Skrzetuski	–	–	1664	0,97%	–	–	1664,01	0,51%
32	Wrzeszczowicz	–	–	1621	0,94%	–	–	1621,009	0,49%
33	Makowiecka	–	–	–	–	1574	2,51%	1574	0,48%
34	Soroka	–	–	1539	0,89%	–	–	1539,009	0,47%
35	Wierszuł	1422	1,51%	91	0,05%	–	–	1513,016	0,46%
36	Stefan Czarniecki	205	0,22%	1273	0,74%	–	–	1478,01	0,45%
37	Kniahini	1447	1,54%	–	–	–	–	1447,015	0,44%
38	Helena	1367	1,45%	34	0,02%	–	–	1401,015	0,43%
39	Miller	–	–	1363	0,79%	–	–	1363,008	0,41%
40	Horpyna	1360	1,45%	–	0,00%	–	–	1360,014	0,41%
41	Sadowski	–	–	1355	0,79%	–	–	1355,008	0,41%
42	Kuklinowski	–	–	1343	0,78%	–	–	1343,008	0,41%
43	Piotr Czarniecki	–	–	1257	0,73%	–	–	1257,007	0,38%
44	Krzysia	–	–	–	–	1190	1,90%	1190	0,36%
45	Kisiel	1108	1,18%	–	–	–	–	1108,012	0,34%
46	Tyzenhaus	10	0,01%	1081	0,63%	–	–	1091,006	0,33%
47	Roch	–	–	1049	0,61%	–	–	1049,006	0,32%

W *Potopie* Zagłoba traci pierwszą pozycję na korzyść Kmicica, którego wysoki wynik jest przynajmniej częściowo rezultatem koncentracji autora na wątku metamorfozy bohatera z libertyna w lojalistę, czemu z kolei służą liczne monologi wewnętrzne, jakie toczy on ze sobą przed, po i w trakcie swej konwersji. Drugie i trzecie miejsca Zagłoby i Wołodyjowskiego wzmacniają ich pozycję jako duetu (stają się jeszcze bardziej nierozłączni), widocznego już zresztą od połowy *Ogniem i mieczem*. W następnej kolejności, podobnie jak w pierwszej części cyklu, pojawia się dwóch głównych antagonistów, a więc obaj Radziwiłłowie<sup>8</sup>. Oczywiście znaczenie wątku publicznego w *Potopie* jest wytłumaczeniem wysokich pozycji dwóch innych „statystów”, Jana Kazimierza i Sapiehy, podczas gdy sąsiadujący z nimi Kordecki wiąże się ze znaczeniem epizodu częstochowskiego w sienkiewiczowskiej wizji potopu szwedzkiego. Główna rola kobieca w tej części cyklu jest, jak wiadomo, znacznie bardziej aktywna: Oleńka musi nie tylko tłumaczyć Kmicicowi, że źle postępuje, nie tylko potem tłumaczyć się ze swego postępowania, ale jeszcze knuć własną uciecz-

<sup>8</sup> Co ciekawe, choć w *Potopie* Bogusław mówi nieco mniej niż Janusz, te kilka słów, które wypowiada w rozdziale IV *Pana Wołodyjowskiego*, przenosi „księcia koniuszego” na wyższą pozycję w rankingu dla całej Trylogii.

kę – nic dziwnego, że w rankingu rozmiarów swych kwestii znajduje się już na 7 miejscu.

W *Panu Wołodyjowskim* całkowitą dominację nad dialogiem przejmuje duet Zagłoba/Wołodyjowski: o ile zyskuje on łącznie prawie 6% całości dialogu w porównaniu ze swym wynikiem z *Ogniem i mieczem*, o tyle proporcje tego duetu są już jednak znacznie bardziej wyrównane wraz z dokonującą się w panu Michale ewolucją od nieodpowiedzialnego młokosa przez „pierwszą szablę Rzeczpospolitej” po „Hektora kamienieckiego” i głównego wyraziciela sienkiewiczowskiego ideału Polaka-żołnierza, ukazanego w ostatniej partii cyklu „w aureoli wielkości”<sup>9</sup>. *Pan Wołodyjowski* jest też jedyną częścią Trylogii, w której „medalowe” miejsce zajmuje postać kobieca. To na pewno nie przypadek, że postać ta ma nawet w pierwszej, czysto romansowej części książki pewne elementy męskie („hajduczek”). Można nawet spekulować, że liczba dzieci, jakiej dochowuje się później główna bohaterka wątku miłosnego każdej z części Trylogii, jest odwrotnie proporcjonalna do liczby wypowiedzianych przez nią słów. Nietypowość tej postaci nie tylko u Sienkiewicza, ale w ogóle na tle polskiej literatury XIX wieku podkreśla jeszcze wypowiedź jednego ze skądinąd najprzychylniejszych krytyków pisarza:

Krzysia podobałaby się nam bardzo, Basia wcale. Nie lubimy panien-kozaków, hajdamaków, źle wychowanych, a przez jakąś dla nich zapewne zrozumiałą kokieterię przesadzających umyślnie złe wychowanie i udających jeszcze gorsze, choć to, które mają, jest już doprawdy dosyć złe. Basia, która wyprawia krzyki i hałasy, [...] przez autora traktowana *con amore*, jest widocznie jego ulubioną kreacją. [...] Nam grunt jej natury i charakteru szlachetny i dzielny podoba się bardzo, ale strona zewnętrzna, układ, sposób mówienia i zachowania z ludźmi daleko mniej. [...] Basia zrobiła zapewne dobrze, że zabiła Tatara [...]; ale zrobiłaby może jeszcze lepiej, gdyby się tym swoim bohaterским czynem nie popisywała, gdyby o nim mówić nie lubiła sama, a nie pozwalała drugim. Jest coś tak przeciwnego naturze, tak wstrętnego w śmierci zadanej ręką kobiety...<sup>10</sup>

Ale w ten sposób Basia wygrywa z Azją nie tylko walkę o własną cześć, ale i o miejsce na liście rankingowej dialogów w trzeciej części Trylogii – choć i tak jego czwarte miejsce jest najlepszym wynikiem osiągniętym przez czarny charakter. W ogóle dalszy układ tej listy wskazuje, że *Pan Wołodyjowski* to znacznie bardziej kameralne dzieło, bo następne w rankingu postacie zawdzięczają swe wysokie pozycje opowiadanym w takim właśnie, kameralnym, świetlicowym kontekście historiom z własnego życia: Muszalski (szczególnie długa opowieść i stąd aż piąte miejsce) i książd Kamiński mają więc do opowiedzenia własne narracje. Polityków reprezentuje tu duet Sobieski/Bogusz, nie ma natomiast wypowiadającej dłuższe kwestie postaci polityka strony przeciwnej – tę rolę musi spełniać Azja. Postaci pani Makowieckiej, Krzysi i nieszczęsnych pani i panny Boskiej są już całkowicie i typowo kobiece.

<sup>9</sup> J. Krzyżanowski, *Twórczość Henryka Sienkiewicza*, Warszawa 1976, s. 149.

<sup>10</sup> S. Tarnowski, „*Pan Wołodyjowski*”, [w:] *Trylogia Henryka Sienkiewicza. Studia, szkice, polemiki*, red. T. Jodełka, Warszawa 1962, s. 224–225.



## Bogactwo słownictwa

Z liczbą słów w tekście wiąże się nierozdzielnie – a wręcz aż za bardzo – zagadnienie bogactwa słownictwa, „bogactwa słownikowego” czy też „gęstości leksykalnej” (w anglojęzycznej literaturze przedmiotu używa się odpowiednio terminów *vocabulary richness*, *lexical richness* i *lexical density*). Za bardzo, ponieważ liczne wskaźniki matematyczne tej i tak dość wieloznacznej cechy stylistycznej tekstu (chętnie stosowanej w atrybucji autorstwa) prędzej czy później okazują się zbyt zależne od wielkości próbki<sup>11</sup> – czyli, w kontekście obecnych rozważań, od tego, czy dana postać mówi dużo, czy mało. Trudności z rozmiarem próbki są zresztą prostym, choć kłopotliwym utrudnieniem w wielu innych badaniach porównawczych idiolektów postaci literackich<sup>12</sup>.

O ile zależność najprostszej, intuicyjnej wręcz miary bogactwa słownictwa, *type-to-token ratio* (TTR), zawarta jest w samym wzorze, o tyle w bardziej skomplikowanych wzorach została stwierdzona stosunkowo niedawno. Jak wykazują w swej cytowanej wyżej pracy Tweedie i Baayen, do badania różnych wielkościowo próbek (czy populacji) nie można nawet bezpiecznie używać owocu myśli jednego z ojców współczesnej statystyki, George’a Udney’ego Yule’a, tzw. liczby *K*, której wzór na pewno nie przekonywał do niej niektórych humanistów:

$$K = \frac{\sum_{r=1}^R r^2 * f(r)}{N^2 - \frac{1}{N}} * 10^4$$

gdzie  $f(r)$  to obserwowana częstość występowania słów (*word-types*) z kolejnych  $r$  miejsc na liście rangowej badanego tekstu ( $r = 1, 2, 3 \dots R$ , gdzie  $R$  to ranga najczęściej występującego słowa),  $N$  to ogólna liczba słów (*word-tokens*) tekstu.

W tej sytuacji orientacyjnie można posłużyć bardzo prostą miarą, wywiezioną wprost z TTR, a stosowaną m.in. w pakiecie oprogramowania Wordsmith Tools: tzw. *standardised type/token ratio* (STTR). Różni się ona od TTR tym, że cały tekst dzielony jest na próbki z góry określonej wielkości (w tym przypadku 1000 słów), po czym następuje uśrednienie wyników danego słowa we wszystkich próbkach.

Dość interesujące wyniki daje lista rangowa 54 głównych postaci Trylogii (postaci pojawiające się w więcej niż jednej części cyklu traktowane są osobno) oparta na wartościach STTR (tabela 6). Dla porównania w tabeli znalazły się też dane dla narracji w każdej z powieści.

<sup>11</sup> F.J. Tweedie, R.H. Baayen, *How Variable May a Constant be? Measures of Lexical Richness in Perspective*, „Computers and the Humanities” 1998, XXXII/5, s. 323–352.

<sup>12</sup> J. Rybicki, *Does Size Matter? A Re-examination of a Time-proven Method*, *Proc. Digital Humanities 2008*, Oulu 2008.

Tab. 6. Bogactwo słownictwa mierzone STTR dla 54 głównych postaci Trylogii i dla narracji\*

Lp.	Postać	STTR	Lp.	Postać	STTR
1	Kordecki	69,24	30	Skrzetuski_1	64,45
2	Kamiński	68,87	31	Wołodyjowski_3	64,44
3	Wierszuł	68,60	32	Bogusz	64,33
4	Narracja_1	68,21	33	Zamoyski	64,20
5	Narracja_2	67,94	34	Nowowiejski	63,87
6	Narracja_3	67,88	35	Chmielnicki	63,62
7	Miller	67,60	36	Zagłoba_1	63,61
8	Jeremi	67,45	37	Piotr_Czarniecki	63,50
9	Sapieha	67,44	38	Kmicic_2	63,46
10	Jan_Kazimierz_2	67,27	39	Wołodyjowski_1	63,05
11	Sobieski	67,04	40	Tyzenhaus_2	63,00
12	Muszalski	66,97	41	Azja	62,66
13	Ketling_3	66,80	42	Oleńka_2	62,24
14	Wrzeszczowicz	66,73	43	Longin	61,97
15	Kisiel	66,70	44	Soroka	61,87
16	Bogusław_2	66,30	45	Kuklinowski	61,60
17	Sadowski	66,10	46	Makowiecka	61,40
18	Sakowicz	66,10	47	Kniahini	61,40
19	Skrzetuski_2	65,90	48	Anusia_2	60,50
20	Czarniecki_2	65,80	49	Rzędzian_2	60,25
21	Zagłoba_2	65,64	50	Kiemlicz	60,16
22	Ketling_2	65,56	51	Helena	59,00
23	Zagłoba_3	65,34	52	Krzysia	58,80
24	Billewicz	65,26	53	Basia	<b>58,70</b>
25	Zaćwilichowski	65,24	54	Rzędzian_1	<b>58,05</b>
26	Wołodyjowski_2	65,17	55	Bohun	<b>57,58</b>
27	Stanisław	65,07	56	Roch	<b>56,70</b>
28	Janusz	64,80	57	Horpyna	<b>56,00</b>
29	Charłamp_2	64,54			

\* Liczby przy nazwiskach niektórych postaci oznaczają część Trylogii. Wartości wytłuszczone są istotne statystycznie (obliczone wg podanego powyżej wzoru)

Już dwa najwyższe wyniki naprowadzają na ciekawy wniosek: dwaj otwierający listę księża wypowiadają u Sienkiewicza długie kwestie, wyraźnie stylizowane na siedemnastowieczną polszczyznę tekstów pisanych lub – w przypadku Kordeckiego – będące wręcz parafrazą autentycznych tekstów postaci historycznej – wszak wiadomo, do jakiego stopnia Sienkiewicz opierał sceny z obłężenia Częstochowy na relacji „świętego przeora”. Przesłanką nakazującą dokładniejsze przyjrzenie się wynikom STTR jest konsekwentnie wysoki i bardzo podobny (trzy sąsiadujące miejsca na liście) wynik języka narracji w każdej z części Trylogii. Obecność w tym towarzystwie Wierszuła, postaci najwyżej drugoplanowej, może dziwić tylko w przypadku, gdy zapomnimy, że znaczną częścią (ponad 75%) jego wypowiedzi w *Ogniem i mieczem* jest stylizowana, epicka opowieść o bitwie pod Piławcami – na co zwraca uwagę już

A. Wilkoń<sup>13</sup>. Stylizowany i narracyjny lub oratorski charakter mają też wypowiedzi następnych postaci w rankingu, postaci dowódców i polityków. Muszalski wypowiada większość swych kwestii w *Panu Wołodyjowskim* w formie gawędy (i dopiero potem zaczyna funkcjonować w powieści jako osobna postać), mało dialogiczne są też wypowiedzi Ketlinga w jego wersji z *Pana Wołodyjowskiego* („Kochanie to niedola ciężka”) i Bogusława Radziwiłła (cyniczne wyznania i knucia, rozmowa z miecznikiem Billewiczem, streszczenia sytuacji politycznej).

Drugą prawidłowością, jaką wyczytać można z tabeli 6 – i to w dodatku jedyłą, co do której można mówić o istotności statystycznej na podstawie wyników testu z) – to fakt, iż na drugim biegunie wyników STTR (wśród najniższych) znajdują się wszystkie postaci kobiece, zajmując miejsca analogiczne do tych, które były ich udziałem w liście rankingowej rozmiarów swych partii. Kobiety u Sienkiewicza mówią więc nie tylko mało, ale i w sposób niezbyt urozmaicony. Można zaryzykować twierdzenie, że w tym męskim świecie są elementem obcym, jednak, jak wynika z omawianej tabeli, nie jedynym, bo w najniższym przedziale gęstości leksykalnej (czy raczej „leksykalno-fleksyjnej”) mieszczą się jeszcze niepolscy i niekatolicy antagoniści z wątków miłosnych powieści, czyli Azja i Bohun (u których wschodnia kwiecistość stylu łączy się z dość monotonnymi wyznaniem miłości i nienawiści), oraz osoby z gminu lub *quasi*-gminu: szczególnie necechowana innością Horpyna (bo nie dość, że kobieta, że Ukrainka, to jeszcze czarownica), Soroka, Kiemlicz i Rzędzian (którego idiolekt w *Potopie* wzbogaca się leksykalnie w porównaniu z *Ogniem i mieczem* wraz ze zbliżaniem się postaci do stanu szlacheckiego). Spostrzeżenie to jest tym cenniejsze, że podobne były czynione innymi metodami statystycznymi na podstawie wyników K. Yule’a<sup>14</sup> i rozkładu częstości najczęściej występujących słów<sup>15</sup>. W tej części tabeli nie brak jednak postaci, których ubogie słownictwo jest widoczne bez jakichkolwiek obliczeń i wynika nawet nie z językowej, ale wprost z opisowej charakterystyki postaci: na pewno należy tu wymienić przede wszystkim Rocha Kowalskiego.

Nieco mniej spodziewanym rezultatem jest znalezienie idiolektów wszystkich pierwszoplanowych postaci, w tym również trzech idiolektów Zagłoby, w środkowej części tabeli. Przecież zarówno najbardziej intuicyjne, impresjonistyczne odczytania Trylogii, jak i najsumienniejsze przeprowadzone badania naukowe – przede wszystkim wymienione już studium Wilkononia – wskazują właśnie na największą oryginalność i zróżnicowanie idiolektu polskiego Falstaffa. Drugim interesującym wynikiem dotyczącym Zagłoby jest nieznaczne rozwarstwienie wyników STTR dla wszystkich trzech wersji tej postaci. Przy nieco większych różnicach i nieco większej ich istotności statystycznej można by mówić wręcz o dwóch idiolektach Zagłoby: mniej zróżnicowanego leksykalnie Zagłoby z *Ogniem i mieczem* i tego drugiego, do którego należą kwestie w dalszych częściach Trylogii. Znacznie mniej zaskakującym efektem jest wyprzedzenie Zagłoby z pierwszej powieści przez Chmielnickiego,

<sup>13</sup> A. Wilkoń, *O języku i stylu „Ogniem i mieczem” Henryka Sienkiewicza. Studia nad tekstem*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego CDXXVII. Prace Językoznawcze”, z. 50, Warszawa-Kraków 1976, s. 36.

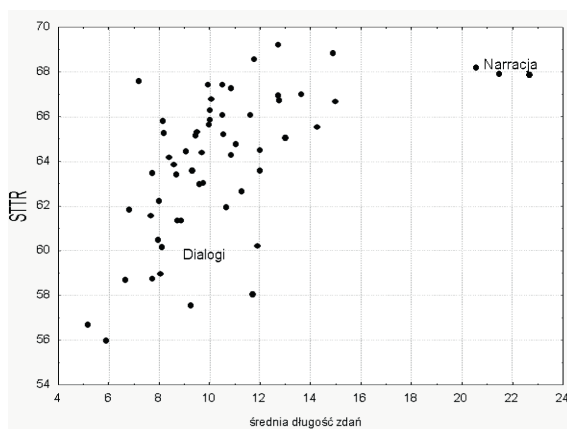
<sup>14</sup> J. Rybicki, *A Computer Assisted Comparative Analysis...*, s. 118–131.

<sup>15</sup> J. Rybicki, *Burrowing into Translation: Character Idiolects in Henryk Sienkiewicz’s Trilogy and its Two English Translations*, „Literary and Linguistic Computing” 2006, XXI/1.

powody tego można bowiem znaleźć w charakterystyce tej drugiej postaci dokonanej przez Wilkonia:

Portret przywódcy kozackiego złożony jest z cech ambiwalentnych. Dialog podporządkowany został koncepcji postaci. W języku Chmielnickiego pojawiają się więc w znacznym zgęszczeniu wymienione wyżej polsko-szlacheckie cechy językowe. Występuje tedy Chmielnicki w dwóch rolach: 1. dwornego szlachcica [...] 2. statysty-oratora. [...] Oratorskie popisy Chmielnickiego przetykane są elementami ukraińskimi i dosadnym słownictwem potocznym<sup>16</sup>.

Podsumowując powyższe i bardziej, i mniej spodziewane wyniki STTR, można powiedzieć, że układają one jednak jednostkowe języki w Trylogii w pewnej logicznej sekwencji: od idiolektów narracyjnych i narracji przez bogate, jednostkowe idiolekty głównych postaci, po ubogie słownictwo bohaterów, u których rzeczywiście należało się tego spodziewać. Co więcej, widoczne w tabeli przemieszanie idiolektów postaci gadatliwych, średnio rozmownych i małowównych to pośredni dowód na słabą zależność uśrednionego TTR od rozmiarów idiolektu – owej największej bolączki znanych miar zróżnicowania słownictwa.



**Wykres 7.** Zależność średniej długości zdań i bogactwa słownictwa w narracji i dialogach w Trylogii

Jak wobec tego odróżnić owe kłopotliwe „narracyjne” idiolekty postaci od samej narracji? Empirycznie bardzo przydatne okazało się połączenie dwóch wymiarów: bogactwa słownictwa mierzonego za pomocą STTR ze średnią długością zdań w narracji i w wypowiedziach poszczególnych postaci (wykres 7). Bez szerszych badań porównawczych nie można stwierdzić, czy taka zależność jest cechą ogólną tekstu literackiego, czy szczególną właściwością stylu pisarskiego Sienkiewicza, czy tylko specyficznym zbiegiem okoliczności w Trylogii. Jednak najbardziej prawdopodobna wydaje się ta pierwsza możliwość: o ile narracja operuje najczęściej długimi zdaniami – niezależnie od rytmicznych lub arytmicznych wahań tej długości – o tyle w najbardziej nawet oratorskiej czy epickiej wypowiedzi postaci zdarzy się krót-

<sup>16</sup> A. Wilkoń, *O języku i stylu „Ogniem i Mieczem”...*, s. 88.

kie, na przykład wykrzyknikowe zdanie. Oto bowiem przeor Kordecki wypowiada niemal jednym tchem zdania długie i krótkie:

Nie wiem zaiste, azali przyjdzie tu zuchwały nieprzyjaciel, ale to wiem, iż gdyby przyszedł, ze wstydem i hańbą odstąpić musi, bo moc jego większa moc pokruszy, złość jego złamie się, potęga startą będzie i odmieni się szczęście jego. Otuchy w serca nabierzcie!

Właśnie na kontraście między zdaniem długim a krótkim oparty jest jego kanzodziejski styl, gdy swe wypowiedzi otwiera apostrofą:

Bracia najmilsi! Niech każdy ku niebu serce podniesie, aby go Duch Święty oświecić raczył, i radźcie potem, mówcie, co któremu sumienie i wzgląd na dobro świętego przybytku dyktuje.

Bracia moi! Nie śpię i ja, gdy wy nie śpicie, modlę się, gdy wy Patronki naszej o ratunek błagacie. Znużenie, trud, słabość czepiają się tak samo kości moich jak waszych; odpowiedzialność tak samo, ba, więcej może na mnie niż na was ciąży – dlaczegoż ja wierzę, a wy już zdajecie się wątpić?...

### Słowa kluczowe

Słowa kluczowe to te, które w sposób statystycznie istotny występują w danym tekście częściej niż w korpusie porównawczym (np. w zbiorze tekstów z analogicznej epoki, w innym tekście tego samego autora, w jakimkolwiek innym tekście). Identyfikacja i rozpoznawanie statystycznej istotności ich frekwencji stała się bardzo prostym zadaniem od chwili pojawienia się programów komputerowych. Oczywiście najwyższą „kluczowość” (*keyness*) osiągają nazwy własne – imiona i nazwiska bohaterów, miejsca akcji; dopiero spośród nich można wyszukać naprawdę ciekawe „klucze” – bardzo często są nimi nie wyrazy „znaczące”, lecz bardziej „mechaniczne” elementy tekstu literackiego: łączniki między narracją a dialogiem, takie jak *rzekł* czy *powiedział*, lub, jeszcze prościej: alternatywne spójniki.

Zagadnienie słów kluczowych jest szczególnie ciekawe w przypadku takim jak Trylogii – gdy chodzi o znalezienie różnic słownikowych między poszczególnymi częściami cyklu. Tabela 7 przedstawia wybraną listę słów kluczowych dla *Ogniem i mieczem* w porównaniu z *Potopem*, po odrzuceniu najbardziej oczywistych, a więc nazw własnych (miejsz i postaci) pojawiających się w pierwszej, ale nie w drugiej części cyklu.

Tabela ukazuje bezwzględną częstość poszczególnych słów w *Ogniem i mieczem* oraz *Potopie*, procentowo częstość względną (udział w ogólnej liczbie *tokens* w danej części Trylogii), stopień „kluczowości” oraz – co najważniejsze – prawdopodobieństwo (P) tego, że dodatni wynik kluczowości jest dziełem przypadku (im niższe, tym wyższa istotność statystyczna; w przedstawionej tabeli wszystkie wartości P są znacznie niższe od najczęściej stosowanej wartości granicznej, czyli 0,05).

Wśród słów, które znalazły się na liście, można wyróżnić kilka grup. Pierwsza zawiera nazwy własne postaci, które co prawda występują również w *Potopie*, ale nie pełnią już równie ważnej funkcji. Nic dziwnego, że należy do nich Skrzetuski, który w drugiej części Trylogii występuje li tylko dla kontynuacji motywu „rycerzy trzech” i nie pełni już funkcji „namiestnika”. Rzędzian z *Potopu* spełnia co prawda na początku swego występu w tej książce rolę łącznika między konfederatami

a Kmicicem, ale na tym jego zadania się kończą. Nawet Zagłoba zdaje się odsuwać nieco od centrum wydarzeń, bo „*pars magna fui*” może powiedzieć o sobie Kmicic (którego odysei Sienkiewicz poświęca znacznie więcej uwagi, miejsca, czasu i narracji bezpośredniej niż w przypadku Skrzetuskiego).

**Tab. 7.** Słowa kluczowe w *Ogniem i mieczem* w porównaniu z *Potopem*

Słowo kluczowe	Częstość w <i>Ogniem i mieczem</i>	%	Częstość w <i>Potopie</i>	%	Kluczowość	P
Skrzetuski	521	0,2178	133	0,0349	464,95	0,0000000000
Skrzetuskiego	207	0,0865	29	–	238,59	0,0000000000
Zagłoba	726	0,3034	561	0,1472	172,55	0,0000000000
Ty	360	0,1505	229	0,0601	125,62	0,0000000000
Jegomość	143	0,0598	43	0,0113	113,67	0,0000000000
Rzędzian	143	0,0598	48	0,0126	104,74	0,0000000000
Ne	61	0,0255	2	–	87,814	0,0000000000
Namiestnika	52	0,0217	0	–	80,264	0,0000000000
Ale	1779	0,7435	2208	0,5794	61,793	0,0000000000
Ja	791	0,3306	880	0,2309	54,017	0,0000000000
Rzeczce	55	0,023	12	–	51,749	0,0000000000
Też	504	0,2106	523	0,1372	47,496	0,0000000000
Która	111	0,0464	62	0,0163	46,772	0,0000000000
Jaru	29	0,0121	0	–	43,639	0,0000000000
Wy	131	0,0548	91	0,0239	38,309	0,0000000000
Zagłoby	86	0,0359	47	0,0123	37,138	0,0000000000
Sława	43	0,018	12	–	34,78	0,0000000008
Rzędziana	40	0,0167	10	–	34,503	0,0000000013
Pohybel	34	0,0142	6	–	34,468	0,0000000014
A	2961	1,2376	4101	1,0761	33,923	0,0000000028
Ku	431	0,1801	473	0,1241	31,326	0,0000000189
Jegomości	55	0,023	24	–	30,859	0,0000000248
Już	1155	0,4827	1484	0,3894	30,019	0,0000000399
Wodę	54	0,0226	24	–	29,677	0,0000000481
I	8923	3,7294	13208	3,4656	29,648	0,0000000489
Zrobił	24	0,01	2	–	29,466	0,0000000540
Rycerz	128	0,0535	102	0,0268	27,63	0,0000001440
Rzplitej	20	–	1	–	26,125	0,0000003171
Gdyż	171	0,0715	156	0,0409	25,441	0,0000004531

Druga grupa to niewątpliwym efektem stosowanej w *Ogniem i mieczem* stylizacji na język ukraiński w dialogach – nawet jeżeli są to słowa równie polskie co „ruskie”, ich częstość wzrasta w wypowiedziach Kozaków: *ja, ty, ne, wy* oraz *pohybel* (choć to ostatnie słowo stosują niemal równie chętnie rdzenni Polacy), również *sława* okazuje się zawdzięczać wysoką frekwencję w pierwszej części cyklu właśnie wypowiedziom kozackim.

Do trzeciej należy zaliczyć te wyrazy, które świadczą o nieco innym rozłożeniu akcentów w słownictwie używanym w archaizacji: w porównaniu z *Ogniem i mieczem* w *Potopie* Sienkiewicz prawie już nie stosuje skrótu „Rzplita”, rzadziej nazywa swe postacie „rycerzami” i używa formy „któren”, z kolei obniżenie częstości względnej zwrotu grzecznościowego „jegomość” to efekt słabszej aktywności dialogicznej Rzędziana w drugiej części Trylogii. „Rzecz” to z kolei dość częsta w pierwszej, a bardzo rzadka w drugiej forma łączenia narracji z dialogiem.

O ile brak jednego choćby jaru w *Potopie* to prosta konsekwencja innego miejsca akcji – a więc moreny i niziny żmudzkie, litewskie i polskie, w górach zaś, przez które przedostaje się do kraju Jan Kazimierz, jest już, podobnie jak w *Pustyni i w puszczy*, „parów” – o tyle nieco bardziej zagadkowa jest kluczowość wody w *Ogniem i mieczem*. Konkordancja dla wykazanej w liście słów kluczowych formy „wodę” wykazuje jednak, że wody rzeczywiście jest sporo w pierwszej części Trylogii. Mamy bowiem i opis posłowania Skrzetuskiego na Sicz – Dnieprem, nad różnymi wodami i rozlewiskami odbywają się bitwy (masakra niemieckich najemników, Żółte Wody, Konstantynów, Beresteczko), potyczki (promowa ucieczka Zagłoby i Heleny) i obłączenia (wyjście Skrzetuskiego ze Zbaraża), Horpyna wróży z wody, „kamień u szyi i w wodę” to też ulubiony kozacki sposób na pozbycie się niepotrzebnych osób.

Jednak pod wieloma względami najciekawsze – bo należące do najbardziej mechanicznych części warsztatu pisarskiego Sienkiewicza – są słowa takie, jak znajdujące się w Tabeli 7 *ku, gdyż, też, a, i* oraz *ale*. W takich przypadkach warto sprawdzić, czy wśród słów kluczowych w odwrotnym porównaniu (tu: między *Potopem* a *Ogniem i mieczem*) nie znajdują się ich inne odpowiedniki.

I rzeczywiście: poza ciekawym faktem, że w *Potopie* występuje statystycznie istotnie mniej konstrukcji ze spójnikami (*a, i*), na pierwszym miejscu pod względem kluczowości napotykamy wyraz *lecz*, będący alternatywnym spójnikiem kontrastującym dla *ale*. Z tego powodu warto przyjrzeć się, jak wygląda rozkład tych dwóch słów kluczowych w poszczególnych częściach Trylogii.

Pakiet Wordsmith Tools umożliwia bezpośrednie wykonanie tego zadania (wykres 8), tworzy bowiem wykres występowania danego słowa w określonych miejscach w tekście (czarne, pionowe kreski).



**Wykres 8.** Rozkład *ale* i *lecz* w trzech częściach Trylogii\*

\* Czarne linie oznaczają wystąpienie danego słowa



Wykres 8 ukazuje ciekawą prawidłowość. Od początku każdej z części Trylogii Sienkiewicz używa przede wszystkim *ale*, i choć w całości tekstu jest to nadal wybór dominujący, to przewaga tego spójnika nad konkurencyjnym *lecz* w początkowych partiach tekstu jest największa w pierwszej, najmniejsza zaś w trzeciej części Trylogii. Potem *lecz* pojawia się coraz częściej, by pod sam koniec każdej z powieści zrównać się lub przewyższyć częstość *ale*. Nie można tu jednak mówić o jakiegokolwiek trwałej ewolucji upodobań Sienkiewicza, ponieważ każda z powieści znów rozpoczyna się od przewagi częstszego z tych dwóch spójników. Najlepiej można zilustrować to przykładami pojawień obu spójników na początku i na końcu pierwszej części Trylogii.

**Tab. 8.** Pierwsze dziesięć wystąpień spójników *ale* i *lecz* w *Ogniem i mieczem*

1	Na łuku Dnieprowym, na Niżu, wrzało jeszcze kozackie życie za porohami, <b>ale</b> w samych Polach nikt nie mieszkał i chyba po brzegach tkwiły gdzieniegdzie „polanki” jakoby wyspy wśród morza.
2	Ziemia była de nomine Rzeczypospolitej, <b>ale</b> pustynna, na której pastwisk Rzeczpospolita Tatarom pozwalała, wszakże gdy Kozacy często bronili, więc to pastwisko było i pobojo-wiskiem zarazem.
3	Step to był pusty i pełny zarazem, cichy i groźny, spokojny i pełen zasadzek, dziki od Dzikich Pól, <b>ale</b> i od dzikich dusz.
4	<b>Ale</b> Tatar, byle wychylił się z Czarnego Lasu lub Dniestr przebył od strony wołoskiej, to stepem równo z ptakami stawał w południowych województwach.
5	<b>Ale</b> blaski gasły coraz bardziej na niebie i na ziemi.
6	Spotkanie pojedynczych cieniów nie znaczyło również nic dobrego, <b>ale</b> nie zawsze należało sobie źle wróżyć, bo i człek żywy zjawiał się nieraz i niknął jak cień przed podróżnymi, dlatego często i snadnie za ducha mógł być poczytany.
7	Żyje, panie namiestniku, <b>ale</b> charcze; arkan go zdławił.
8	<b>Ale</b> koń, panie namiestniku, będzie nasz? wtrącił tonem pytania wachmistrz.
9	W oczach jego malowała się okrutna fantazja i zadzierzystość, <b>ale</b> w obliczu miał wyraz uczciwy.
10	<b>Ale</b> wybacz waszmość pan, zem mu naprzdó powinien nie złożył dzięki za auxilium i skuteczny ratunek, który mnie od tak nagłej śmierci wybawił.

**Tab. 9.** Ostatnie dziesięć wystąpień spójników *ale* i *lecz* w *Ogniem i mieczem*

1846	<b>Ale</b> najodważniejszym wydawało się to niepodobieństwem.
1847	Bronić się w nim można było całe lata, <b>ale</b> do odwrotu jedna tylko droga stała otworem – przez wojska królewskie.
1848	<b>Lecz</b> wkrótce przekonał się, że z tymi zastępami nie można było już myśleć o przejściu zbrojną ręką po trupach królewskiego wojska – więc chwycił się innego sposobu.
1849	Król zwłóczył szturm nie chcąc przelewu krwi, <b>lecz</b> widząc te olbrzymie roboty poznał, iż nie ma innej rady, i kazał otrąbić w wojsku, by się na wieczór gotowano do ostatecznej rozprawy.
1850	<b>Lecz</b> na próżno zawrócił w tej chwili do taboru, na próżno leciał naprzeciw tłumom ze wzniesionymi do nieba rękami.
1851	Wojska koronne zdumiały się na widok tego ruchu, który zrazu wielu za jakiś atak rozpaczliwy poczytało; <b>lecz</b> oczom trudno było nie wierzyć.
1852	Nie śpiewano <i>Te Deum</i> i nie łyż radości, <b>lecz</b> łyż żalu i smutku płynęły z dostojnych oczu królewskich.
1853	<b>Lecz</b> Bohun nie złożył wraz z innymi głowy w tym dniu straszliwym.
1854	Postrzał z jakiejś mściwej ręki dosięgnął go w kilka lat później, <b>lecz</b> i wówczas nie przyszedł na niego kres ostatni.
1855	Żył nie w łubniach, <b>ale</b> w wiosce, którą z popiołów odbudował i która zwała się Rozłogi.

## Podsumowanie

Jak stwierdzono we wstępie, z liczenia i statystyki tekstu literackiego możemy spodziewać się dwójakiego rodzaju korzyści. Po pierwsze policzenia podstawowych elementów i zjawisk – od kiedy szacunkowe określanie średniej liczby słów w każdym wierszu książkowej strony i mnożenie ich przez liczbę wierszy zostało wyparte przez znacznie dokładniejsze programy do edycji tekstu – daje w prosty sposób wyniki dokładne, niezmiennie i wiarygodne. Jest to oczywiście bardzo podstawowy opis rzeczywistości, który do prawdziwej, naukowej pracy nad tekstem ma się mniej więcej tak jak w badaniach przyrodniczych liczenie ptaków na danym obszarze do eksperymentów genetycznych, a jednak zoologowie do dziś liczą ptaki i inne stworzenia. Przyjrzenie się rozmiarom rozdziałów i zdań to już nawet dla najmniej ceniącego arytmetykę humanisty całkiem sensowna działalność, bo wielkości te od dawna uznawane są za cechy stylistyczne. Bo o ile nie potrzebujemy komputera, by zauważyć, że długości i struktura zdań to najprostszy sposób odróżnienia stylu Prousta od stylu Hemingwaya, to znalezienie istotnych statystycznie różnic między tymi samymi wartościami dla autorów o mniej charakterystycznym stylu czy dla różnych tekstów tego samego autora jest cennym odkryciem (i tylko trochę cenniejszym od odkrycia, że różnic tych brak np. między autorami tego samego okresu literackiego, tej samej tendencji, tego samego ugrupowania artystycznego). Ewolucje stylu obserwowane właśnie na policzalnych cechach pisarstwa są już bardzo dobrze opisane w literaturze przedmiotu, i to zarówno dla indywidualnych pisarzy, jak i dla całych literatur narodowych<sup>17</sup>.

Drugą ze wspomnianych korzyści jest odkrywanie niewidocznych owym gołym okiem zjawisk czy to dzięki prostym zabiegom rachunkowym, czy też nieco bardziej wymyślnymi metodami, jak choćby określania bogactwa słownikowego czy kluczowości słów w tekście literackim. Badanie idiolektów postaci to już niemal klasyczna pozycja w repertuarze tej gałęzi literaturoznawstwa: śledzenie istotnych statystycznie zmian upodobań autorskich co do poszczególnych słów to już znacznie więcej niż proste rachunki.

Oczywiście literaturoznawstwo komputerowe (bo tak chyba należy tłumaczyć okrzepły już angielski termin *literary computing*) stosuje znacznie bardziej wymyślne, profesjonalne, efektywne i efektowne techniki obliczeniowe i statystyczne: analiza głównych składowych, Delta Burrowsa, analiza dyskryminacyjna, fourierowska analiza sekwencyjna, sieci neuronowe – to wszystko jest stosowane od wielu lat i z dobrym skutkiem; główną bolączką tego środowiska naukowego pozostaje dylemat, jak z niewątpliwie ciekawymi wynikami zastosowań „wyższej statystyki” w badaniach literackich dotrzeć do szerszej literaturoznawczej publiczności. Jednym z rozwiązań mogą być właśnie takie proste rekonesanse – ukazanie, że opisanie najprostszych policzalnych zjawisk i cech tekstu literackiego jest nie celem samym w sobie, lecz nowym rodzajem wstępnego działania w służbie refleksji czysto literaturoznawczej. Jak obrazowo stwierdza w jednym ze swych najnowszych tekstów cytowany już David Hoover:

---

<sup>17</sup> Np. D. Hoover, *Corpus Stylistics, Stylometry, and the Styles of Henry James*, „Style” 2007, XLI/2, s. 174–203, lub J.F. Burrows, *Tiptoeing into the Infinite: Testing for Evidence of National Differences in the Language of English Narrative*, „Research in Humanities Computing” 1994, II, s. 1–33.

Odchodzimy od tekstu, kryjemy się w gąszcz liczb, by bogatsi o zgromadzoną w ten sposób wiedzę do tekstu powrócić. Bez zrozumienia tekstu nie zrozumiemy liczb; ale liczby pomogą nam uchwycić nieuświadomiane dotąd sensy tekstu, które z kolei mogą stać się lekarstwem na postmodernistyczny relatywizm poznawczy współczesnej humanistyki i jej brak wiary w możliwość wypracowania „naukowej” czy „obiektywnej” metodologii<sup>18</sup>.

## Counting Sienkiewicz's Trilogy

### Abstract

This paper is an attempt to show how various countable features of a literary text – exemplified here by the trilogy of historical romances by Henryk Sienkiewicz – can be helpful in contrastive analyses between writings by two or more authors, writings by the same author, or even different fragments of the same text (volumes, chapters, characters' individual languages). The simplest measures (leading to the simplest conclusions) include text size expressed in various units (number of words, sentences, paragraphs) and proportions in various modes of writing (e.g. dialogue vs. narration); somewhat more complex is the calculation and the evaluation of vocabulary richness (or lexical density) – the paper proposes Standardized Type/Token Ratio as the method exhibiting the least dependence on text size. At this point, the importance of calculating statistical significance is claimed.

Distribution graphs are useful in presenting the trends in chapter length, which behave differently for different parts of the Trilogy. Sentence length is confronted with content; the longest sentences in Sienkiewicz's masterpiece have been found to be connected with the political/military context.

The sizes of individual characters' parts in the dialogue might be found in interesting connection with the significance of each character in the story; in Sienkiewicz, feminine parts are markedly shorter than masculine. The same is true of vocabulary richness of these parts. The study of keywords (words statistically significantly more or less frequent in one text as compared with another text) also yields interesting information on Sienkiewicz's writing technique. Most importantly, a repeated phenomenon has been observed: in all parts of the Trilogy, the initially infrequent *lec* tends to gradually increase its occurrence at the expense of the generally more frequent *ale*.

---

<sup>18</sup> D. Hoover, *Quantitative Analysis and Literary Studies*, [w:] *A Companion to Digital Literary Studies*, red. S. Schreiban, R. Siemens, Oxford 2008, s. 521 (tłum. J.R.).