

Adam Pawłowski

Uniwersytet Wrocławski
Wrocław

Maciej Eder

Uniwersytet Pedagogiczny im. Komisji Edukacji Narodowej
Kraków
Instytut Języka Polskiego PAN
Kraków

Struktury sekwencyjne w *Kronice Dalimila*: analiza stylometryczna¹

1. Przedmiot i cel badań

Analizując metodami stylometrycznymi staroczeską *Kronikę Dalimila* (*Staročeská kronika tak řečeného Dalimila*), Jerzy Woronczak dostrzegł w kolejnych jej rozdziałach proces powolnego rozpadu regularnej struktury wiersza (Woronczak 1993 [1963]). Zjawisko to określił mianem prozaizacji. W istocie zauważyć można, że w kolejnych częściach *Kroniki*, w miarę jak zmniejsza się historyczny dystans, dzielący opisywane wydarzenia od czasów współczesnych autorowi i czytelnikowi (lub słuchaczowi), faktura wersyfikacyjna staje się mniej regularna – zaczyna przypominać prozę zapisaną w formie wersów. W celu wyjaśnienia swoich spostrzeżeń Woronczak sformułował hipotezę, zgodnie z którą wierszowaną narra-

¹ Niniejszy artykuł jest zmienioną i rozwiniętą wersją tekstu pochodzącego z tomu *Sequences in Language and Text* (Pawłowski, Eder 2015).

cję historiograficzną konstruuje się łatwiej, gdy opisywane zdarzenia stanowiące jej kanwę leżą poza zasięgiem pamięci współczesnych autorowi odbiorców – na przykład w czasach biblijnych lub przedchrześcijańskich. Natomiast przy opisie wydarzeń, których naoczni świadkowie żyją bądź też istnieją na ich temat wiarygodne przekazy ustne, materia języka stawia większy opór. O ile więc początkowe fragmenty *Kroniki* są bardziej regularne, ponieważ można uznać je za wykład mitologii narodowej z elementami historiografii, w jej końcowych częściach sytuacja jest odwrotna. Dlatego kronikarz nie mógł pozwolić sobie na dostosowywanie treści do ograniczeń wersyfikacyjno-metrycznych i rymowych, jakie narzucała przyjęta forma narracji wierszowanej – opis musiał być wierny, a więc zgodny z oczekiwaniami i wiedzą potencjalnych odbiorców z kręgu dworskiego².

Można wysunąć dodatkowy argument przemawiający za taką tezę. Początkowe partie *Kroniki Dalimila* z dużym prawdopodobieństwem zawierają ślady tradycji oralnej, opierającej się na monotonnym powtarzaniu tych samych fraz, tworzących długie konstrukcje parataktyczne (por. Havelock 2006). Struktury takie są stosunkowo łatwe do zapamiętania i najprawdopodobniej odpowiadały mnemotechnicznym umiejętnościom człowieka przedpiśmiennego. U schyłku średniowiecza, kiedy w europejskiej przestrzeni komunikacyjnej rozpoczynała się epoka dominacji pisma, formulaiczne teksty literatury oralnej stopniowo znikwały z obiegu kultury. Niektóre z nich trafiały jednak do nowego medium w postaci zapożyczeń międzystylowych³. Jest więc prawdopodobne, że początkowe fragmenty *Kroniki* zawierają fragmenty o wiele starszych tekstów, należących do tradycji oralnej, co sprawia, że wersyfikacja początkowych partii *Kroniki* jest bardziej regularna, niż rozdziałów końcowych.

Pomysł Jerzego Woroncza był tyleż prosty, co uniwersalny, ponieważ pozwalał lepiej zrozumieć utrwalone w kulturze trudne współistnienie fikcyjności i oczekiwań czytelniczych w tym zakresie (a można przyjąć za pewnik, że odbiorcy treści symbolicznych w każdej epoce i kulturze chętnie akceptują atrakcyjną fikcję) z potrzebą prawdy jako kategorii opisu historiograficznego⁴.

² Prof. Woroncza przedstawił to wyjaśnienie podczas jednej z dyskusji na temat metod ilościowych w badaniach filologicznych, jaka w latach dziewięćdziesiątych XX wieku odbyła się w Jego gabinecie w Instytucie Filologii Polskiej Uniwersytetu Wrocławskiego. Słuchaczem Jego wywodów na ten temat i w skromnym zakresie rozmówcą był Adam Pawłowski, wówczas adiunkt w IFP UW. Według najlepszej wiedzy autorów niniejszego artykułu wyjaśnienie to nie zostało nigdy przez prof. Woroncza upublicznione.

³ Stosunkowo wyraziście proces tej swoistej konwersji na nowe medium widać w tekście *Starego Testamentu*, gdzie istnieją liczne zapożyczenia z literatury oralnej (por. Cullery 1986).

⁴ Nie wnikając w poważne rozważania epistemologiczne, przyjęto, że „prawda” jest kategorią oznaczającą intersubiektywnie odczuwaną zgodność komunikatów tekstowych z doświadczeniem podmiotu poznającego.

Jego podejście metodologiczne było przekonujące, ponieważ opierało się na analizie materiału empirycznego *Kroniki* i wykorzystaniu narzędzi statystyki. Jednak Woronczak prowadził swoje badania w latach sześćdziesiątych XX wieku i nie mógł pracować na tekście elektronicznym, pozwalającym na pełne wykorzystanie narzędzi matematycznych i informatycznych. Zastosował jedynie tzw. test serii, pozwalający na liczbowe wyrażenie stopnia losowości szeregu dowolnych obserwacji. Zaletą takiego podejścia jest ujmowanie materiału empirycznego jako sekwencji danych (kolejność jednostek jako cecha relewantna), a nie zbioru (kolejność jednostek nierelewantna) oraz możliwość wykorzystania stosunkowo łatwych do pozyskania z tekstu danych liczbowych. O wiele skuteczniejsze w tym kontekście są jednak inne metody, a mianowicie analiza szeregów czasowych, pozwalająca na rozpoznanie i matematyczne modelowanie stabilnych tendencji zmian różnych parametrów stylometrycznych tekstu, a także powtarzających się wzorców rytmicznych⁵.

Celem niniejszej pracy jest weryfikacja niektórych założeń Jerzego Woronczaka za pomocą narzędzi analizy sekwencyjnej. W szczególności analizie poddane zostaną: (1) szeregi utworzone przez długości rozdziałów (obliczone w literach, sylabach i słowoformach); (2) szeregi utworzone przez długości wersów (liczone w literach, sylabach i słowoformach); (3) alternacje i korelacje par rymowych; (4) szeregi iloczynowe (sylaby długie i krótkie, kodowanie binarne); (5) szeregi akcentowe (sylaby akcentowane i nieakcentowane, kodowanie binarne). Stosowana metoda była już wcześniej z powodzeniem testowana w pracach na temat rytmiki tekstów w językach nowożytnych (francuski, polski, rosyjski) i klasycznych (łacina i greka)⁶.

Posługując się materiałem empirycznym, zamierzamy zbadać wzorce rytmiczne w tekście *Kroniki* i w ten sposób poddać weryfikacji hipotezę Woronczaka. Spodziewane rezultaty dotyczą w pierwszej kolejności samej *Kroniki Dalimila*. Jednak dyskusja nad oralnością i piśmiennością – historycznie dwiema podstawowymi formami interakcji werbalnej – wykracza poza filologiczne kwestie średniowiecznej annalistyki i ma uniwersalny charakter.

⁵ Test serii oraz podstawy metody szeregów czasowych zostały szczegółowo omówione w rozdziale *Metoda badań* i w *Aneksie*.

⁶ Por. Pa w ł o w s k i 1998, 2003; Pa w ł o w s k i, E d e r 2001; Pa w ł o w s k i et al. 2010; E d e r 2008.

2. Charakterystyka *Kroniki*

Kronika tak zwanego Dalimila (Staročeská kronika tak řečeného Dalimila) jest jednym z najstarszych i z pewnością najważniejszym zabytkiem średniowiecznej czeszczyzny, należącym do zbioru mitów założycielskich Królestwa Czech. Powstała na początku XIV wieku, prawdopodobnie jako dzieło anonimowe, a potomność nadała jej nieznanemu autorowi (lub autorom) imię Dalimila. Dzieło, napisane nieregularnym wierszem rymowanym, zawiera części oryginalne, będące translacją fragmentów starszych teksów (np. *Chronica bohemica*) oraz partie dopisane w późniejszych okresach. Od chwili powstania (a raczej odkrycia przez filologów) miała trzy krytyczne wydania: pierwsze powstało w roku 1882 (Jireček 1882), drugie w 1957 (Havránek, Daňhelka 1957), trzecie w 1988 (Daňhelka et al. 1988). W omawianych tutaj badaniach wykorzystano wersje z 1882 i 1988 roku. Ponieważ segmentacje *Kroniki* w poszczególnych rękopisach nieznacznie się różnią, jako podstawę obliczeń przyjęto numerację z wydania Jirečka (1882), obejmującą sto sześć rozdziałów (cz. *kapitola*) z przedmową⁷.

Kronika Dalimila opisuje historię Królestwa Czech od jego mitycznych początków (konstrukcja Wieży Babel) do roku 1314. Linearny porządek rozdziałów odpowiada chronologicznemu następstwu zdarzeń, nie ma jednak dowodów na to, że proces tworzenia *Kroniki* przebiegał także linearnie i pokrywał się z kolejnością przedstawianych epizodów. Jest prawdopodobne, że pewne fragmenty mogły powstać jako uzupełnienie części już istniejących, inne natomiast były przerezegowane.

Warto także pamiętać o niejasnościach związanych z autorstwem *Kroniki Dalimila*. Nie można wykluczyć, że autorów tego dzieła było kilku, a jeżeli pozostanie się przy hipotezie, mówiącej o istnieniu jednego twórcy, nie wiadomo, jakie fragmenty są jego/jej oryginalnym wkładem, a jakie są zapisem innych tekstów funkcjonujących w ówczesnym obiegu oralnym i rękopiśmienniczym. Wątpliwości nie budzi jedynie to, że wydarzenia przedstawione w początkowych częściach *Kroniki* leżały całkowicie poza zasięgiem pamięci autora i jemu współczesnych (miały więc charakter częściowo fikcyjalny), a w miarę postępu narracji oś czasu świata przedstawionego zbliżyła się do okresu, którego znajomość mogła pochodzić z autopsji lub pośrednich przekazów świadków zdarzeń. Jest także bardzo prawdopodobne, że wiedza „tak zwanego Dalimila” pochodziła z wcześniejszych źródeł łacińskich lub z literatury popularnej, przekazującej z pokolenia na pokolenie fabularyzowane opowieści. Świadczy o tym przede wszystkim typowa dla literatury oralnej redundancja tekstu, wynikająca z powtarzalności fraz, stałych epitetów, paralelizmu składniowego, anaforycznych początków wersów itd.

⁷ W innych wersjach *Kroniki* występują sto trzy rozdziały.

Mimo powyższych zastrzeżeń wkomponowanie w tekst *Kroniki Dalimila* fragmentów wcześniejszych utworów, a także późniejsze zmiany redakcyjne, nie są przeszkodą w prowadzeniu analizy sekwencyjnej tego dzieła jako całości. Przy pewnym poziomie wewnętrznego zróżnicowania *Kronika* pozostaje bowiem tekstem spójnym tematycznie i formalnie, w którym narracja i fabuła rozwijają się równolegle w porządku linearnym.

3. Hipoteza

Wysunięto hipotezę, zgodnie z którą w tekście *Kroniki* można zaobserwować stopniową zmianę faktury wersyfikacyjnej, polegającą na rosnącym zróżnicowaniu (dysymilacji) przylegających do siebie wersów. Proces ten, określony przez Jerzego Woronczaka mianem prozaizacji (Woronczak 1993 [1963]: 70), jest prawdopodobnie efektem zmiany preferencji stylistycznych autora, a także jego/jej stosunku do faktów historycznych. Może jednak również odzwierciedlać przejście od tradycji literatury oralnej, opartej na bardzo regularnej i prostej strukturze rytmicznej, której funkcją było ułatwianie zapamiętywania i publicznego odtwarzania tekstu, do „nowoczesnej” kultury pisma, pozwalającej na tworzenie złożonych i wyszukanych wzorców rytmicznych, odpowiadających bardziej wyrafinowanym gustom odbiorców dworskich. Oprócz weryfikacji powyższej hipotezy, wynikającej jasno z rozważań Woronczaka, przeprowadzono dodatkowe analizy, których celem było poszukiwanie regularnych struktur, tworzonych przez sekwencje sylab akcentowanych dynamicznie i iloczynowo. Jednym z tego powodów jest fakt, iż analogiczne podejście w odniesieniu do tekstów pisanych greckim i łacińskim heksametrem dało wartościowe wyniki (por. Pawłowski, Eder 2001; Pawłowski et al. 2010).

4. Metoda

Ilościowa analiza tekstu jest kombinacją matematyki stosowanej i lingwistyki. Wykorzystanie metodologii ścisłej nadaje takiemu podejściu charakter naukowy, jednak przy bliższej analizie może budzić pewne wątpliwości, o których warto tutaj wspomnieć. Złożoność i wieloaspektowość języka są w badaniach ilościowych redukowane do niewielkiej liczby kwantyfikowalnych parametrów, a aspekty semantyczne i pragmatyczne, czyli esencja języka i komunikacji – są traktowane wciąż jeszcze powierzchownie. Warto też zauważyć, że w praktyce badawczej często brakuje krytycznej weryfikacji uzyskanych wyników, prowadzonej z perspektywy jakościowej. Mimo tych zastrzeżeń metody ilościowe okazały się sku-

teczne w rozwiązywaniu wielu kwestii praktycznych (choćby w stenografii, leksykografii czy glottodydaktyce, a współcześnie również w badaniach komunikacji elektronicznej), przyczyniły się także do rozwoju badań podstawowych, na przykład w ramach tzw. lingwistyki synergetycznej, ujmującej w postaci modeli matematycznych relacje pomiędzy różnymi zmiennymi, charakteryzującymi teksty i słownictwo. Osiągnięcia te mają swój początek w badaniach dziewiętnastowiecznych, ale ich rozkwit był możliwy dopiero w epoce cyfrowej, dzięki zastosowaniu narzędzi obliczeniowych do wielkich korpusów tekstów. Te nowe możliwości znacznie poszerzyły, przynajmniej w sensie ilościowym, perspektywę myślenia o języku. Jak zauważa Reinhard Köhler:

W ostatnich latach odkrytych zostało wiele praw językowych, a niektóre spośród nich zintegrowano z ogólnym modelem synergetycznym. Z tego względu lingwistyka synergetyczna może być uważana za pierwszy załamek teorii lingwistycznej. [...] Zgodnie z zasadami filozofii nauki istnieje jeden, powszechnie uznawany schemat wyjaśniania: dedukcyjno-nomologiczny (Köhler 2005: 764, przekład własny).

Uznanie lingwistyki synergetycznej za „pierwszy załamek teorii lingwistycznej” jest chyba pewnym nadużyciem, ale bez wątplenia paradygmat systemowy (a więc odwołujący się do ogólnej teorii systemów, uznający język za system podlegający prawom ogólnym), połączony z możliwościami narzędzi analizy ilościowej, umożliwia o wiele głębsze zrozumienie zagadnień filologicznych.

Niniejsze studium powstało dzięki połączeniu aparatu filologicznego z ilościowymi technikami modelowania sekwencyjnego. Przedstawione dalej wyniki wnoszą przede wszystkim nową wiedzę na temat *Kroniki Dalimila* – wyjątkowego zabytku literatury i kultury staroczeskiej. Ponadto część refleksji dotyczy kwestii ogólniejszych, a przy tym ważnych dla badań filologicznych i teoretycznoliterackich, takich jak relacja między faktografią i fikcjonalnością w dawnej historiografii oraz proces przejścia od oralności do piśmienności w średniowiecznej Europie. Warto podkreślić, że przedstawione dalej badania zostały oparte na empirycznych i ilościowych analizach średniowiecznej wersyfikacji, a tylko w niezbędnym zakresie, wyznaczonym dostępnością danych, na wnioskowaniu indukcyjnym i uprawdopodobnionej spekulacji (dotyczy to chociażby uwag na temat hipotetycznych oczekiwań czytelniczych epoki średniowiecza, wymuszających określone praktyki dyskursywne).

Analiza ilościowa wersyfikacji wykorzystuje metody statystyki konwencjonalnej oraz narzędzia stosowane w badaniach struktur sekwencyjnych. Od publikacji *Language as a Choice and Chance*, monumentalnej monografii autorstwa brytyjskiego fizyka i lingwisty Gustava Herdana, znane z literatury przedmiotu użycia statystyki zakładają postrzeganie języka albo jako zbioru (systemu) ele-

mentów (*analysis of language in the mass*), albo jako sekwencji tychże (*analysis of language in the line*) (Herdan 1966: 423). Głównymi narzędziami analizy sekwencyjnej, opierającej się na założeniu, zgodnie z którym relewantną cechą języka jest liniowy porządek elementów w tekście, są metoda szeregów czasowych (prowadzona w dziedzinie czasu lub częstotliwości), teoria informacji oraz różne teorie procesów stochastycznych. Wszystkie one były wykorzystywane w dotychczasowych badaniach, ale metoda szeregów czasowych okazała się najbardziej skuteczna i właściwa dla analizy tekstów (Pawłowski 1998, 2001).

Koncepcja analizy sekwencyjnej wychodzi od elementarnej opozycji stosunków syntagmatycznych i paradygmatycznych w tekście, stanowiąc jej rozwinięcie. Aby wyjaśnić jej istotę, warto posłużyć się przykładem. Przyjmijmy, że w hipotetycznym języku *L* istnieją dwa typy jednostek, oznaczanych jako A i B, tworzących, zgodnie z zasadami składni, struktury linearne, zwane tekstami. Przykładowy minikorpus takich tekstów można przedstawić jako sekwencje:

- (1) AAAAAABBBBBB, (2) AAABBBAAABBB, (3) AABBAABBAABB,
 (4) ABABABABABAB, (5) AABBABABABAB, (6) BABBAABBBAAA

Z punktu widzenia lingwistyki i komunikacji sekwencje te różnią się między sobą w taki sam sposób, jak wyrazy *czapka* i *paczka*, zawierające dokładnie ten sam zestaw liter i/lub fonemów. Metody statystyki konwencjonalnej, oparte na obliczaniu parametrów pozycyjnych i rozkładów statystycznych, nie wykażą różnic między nimi, ponieważ częstości jednostek A i B są identyczne w każdym „tekście” (różnica wystąpi jednak przy uwzględnieniu częstości bigramów i trigramów). W takim przypadku analiza sekwencyjna, uwzględniająca porządek jednostek składowych, jest o wiele skuteczniejszą metodą badawczą. Oczywiście dłuższe teksty w języku naturalnym rzadko kiedy prezentują identyczne rozkłady jednostek, szczególnie w sytuacji, gdy analizuje się leksykę lub te własności, które przyjmują bardzo zróżnicowane wartości. Ale w analizie metryki (a tutaj mamy do czynienia z taką sytuacją) przy kwantyfikacji tekstu stosuje się z powodzeniem skale dwuwartościowe, czyli binarne kodowania iloczasu lub akcentu dynamicznego. Wówczas znalezienie różnych tekstów o identycznej lub podobnej liczbie jednostek nacechowanych i nienacechowanych jest już prawdopodobne.

Należy podkreślić, że struktury syntagmatyczne, określane niekiedy również jako *szeregi tekstowe* (przez analogię do szeregów czasowych), nie są jedynymi konstruktami linearnymi, jakimi zajmuje się nauka o języku, i jakie można modelować przy pomocy narzędzi matematycznych. Jako szeregi czasowe traktuje się również sekwencje pełnych tekstów lub ich fragmentów. Dobrym tego przykładem są dzieła jednego autora, uporządkowane zgodnie z chronologią ich powstania (Lutosławski 1897a), periodycznie ukazujące się artykuły prasowe lub

uporządkowane chronologicznie próbki dyskursu publicznego (Pawłowski 2006). Ponieważ kwantyfikacja przebiega w takich przypadkach najczęściej (ale nie wyłącznie) w oparciu o frekwencje jednostek leksykalnych, można tutaj mówić o *szeregach leksykalnych*. Warto przypomnieć, że cytowany tutaj Wincenty Lutosławski wprowadził do nauki o języku pojęcie *stylemu*, które oparte było właśnie na frekwencjach leksemów⁸.

Istnieje kilka metod matematycznych, pozwalających na analizę sekwencji. Należą do nich teoria informacji z pojęciami entropii i redundancji, metoda łańcuchów Markowa, analiza spektralna. W tym przypadku wykorzystano analizę szeregów czasowych. Obejmuje ona wiele komponentów, takich jak funkcja autokorelacji (ACF) i autokorelacji częściowej (PACF), a także proste i złożone modele procesów stochastycznych, takie jak model autoregresji rzędu $AR(p)$, model ruchomej średniej $MA(q)$ oraz ich kombinacje (modele ARMA). Pełny opis szeregu czasowego w dziedzinie czasu obejmuje trzy składowe: periodyczną (reprezentowaną przez funkcję i model autoregresji i/lub ruchomej średniej), losową (odpowiadającą kategorii szumu lub zakłóceń procesu) oraz trend, czyli deterministyczny element o charakterze nieperiodycznym, dający się wyrazić jakąś funkcją monotoniczną (tutaj zostanie zastosowana funkcja liniowa).

Z punktu widzenia lingwistyki rząd modelu (p lub q) odpowiada głębokości statystycznie istotnego związku z lewostronnym kontekstem. Można więc tutaj mówić o głębokości „pamięci tekstowej”, czyli determinowaniu danego elementu sekwencji przez p lub q jednostek bezpośrednio poprzedzających. Ponieważ artykuł niniejszy nie ma charakteru metodologicznego, a zasady analizy szeregów czasowych zostały szczegółowo opisane w literaturze przedmiotu, a w szczególności w pracach przeznaczonych dla ekonomistów i inżynierów (por. Box, Jenkins 1976; Cryer 1986), przedstawicieli nauk społecznych (McCleary, Hay 1980), psychologii (Glass, Wilson, Gottman 1975; Gottman 1981, Gregson 1983, Gottman 1990) oraz lingwistyki (Pawłowski 1998, 2001, 2005; Pawłowski, Eder 2001; Pawłowski et al. 2010), ich szczegółowego opisu nie przedstawiamy, ograniczając się do podania podstawowych informacji w *Aneksie*.

Warto jednak pamiętać o tym, że sekwencja dyskretnej obserwacji zmiennej losowej zwana jest szeregiem czasowym, jeżeli może być w sensowny spo-

⁸ W fundamentalnej pracy Lutosławskiego *The origin and growth of Plato's logic* (Lutosławski 1897a: 145–155) stosowane jest jeszcze synonimiczne wyrażenie *stylistic peculiarity*, ale termin *styleme* pojawia się w innych jego publikacjach już w tym samym roku (por. Lutosławski 1897b). Warto przytoczyć definicję stylemu Lutosławskiego, ponieważ mimo pewnej nieostrości zawiera ona w sobie główny pierwiastek definicji statystycznej stylu: „Un styleme est tout usage du langage qui n'est pas indispensable pour exprimer la pensée, c'est-à-dire, qui peut être remplacé par un autre” (Lutosławski 1898: 65).

sób dopasowana do osi reprezentującej czas lub inną, porównywalną, wielkość. Dowolny szereg czasowy może być rozłożony na trzy składowe: trend, funkcje periodyczne (w tym także oscylacje sezonowe, pojawiające się w nieciągłych odstępach) oraz tzw. biały szum⁹. W lingwistyce takie „pełne” szeregi w zasadzie nie występują. W przypadku analizy pojedynczego tekstu nie pojawia się trend, ponieważ dane są stacjonarne, a więc stabilne w pewnym przedziale, wyznaczonym minimalną i maksymalną frekwencją¹⁰. Istnienie trendu dla wybranego parametru i jego estymacja są jednak możliwe przy analizie szeregu zbudowanego z wielu tekstów, uporządkowanych chronologicznie lub według innej zasady. W taki sposób (choć bez narzędzi analizy chronologicznej) badania tekstów Platona prowadził na przykład Wincenty Lutosławski (1897a), metodę taką stosujemy również podczas badania kolejnych segmentów *Kroniki Dalimila* ze względu na średnią długość wersu. Jednak nawet tego rodzaju szeregi tekstowe zachowują pewną stacjonarność, ponieważ wartości używanych parametrów nie są nieograniczone.

Spośród wszystkich parametrów, jakie wykorzystuje się przy analizie szeregów czasowych lub innych danych poziomych, w których można spodziewać się oscylacji periodycznych (czyli jakiejś formy rytmiczności), najważniejsza jest funkcja autokorelacji szeregu. Jeżeli jej wartości są niskie i nieznaczące, nie ma potrzeby wykonywania kolejnych kroków analizy, ponieważ uzyskane wyniki będą także nieznaczące.

5. Test serii w badaniach Jerzego Woronczaka

W analizie będącej punktem wyjścia naszych badań Jerzy Woronczak zastosował nieparametryczny test serii, wykorzystywany do oceny stopnia losowości szeregów reprezentujących sekwencje wyników obserwacji lub eksperymentów (Woronczak 1993 [1963]). Aby dostrzec zalety i wady tego testu, można posłużyć się przykładem. Wyobraźmy sobie sekwencję symboli A i B {AABBA-AAABBABBBAAABBABB}, gdzie n_a jest liczbą jednostek A, n_b jest liczbą jednostek B, natomiast N jest długością serii ($n_a + n_b = N$). Jeżeli prawdopodobieństwa wystąpienia jednostek A i B są znane, możliwa jest estymacja rozkładu przylegających do siebie wystąpień A i B (czyli właśnie serii) oraz przedziałów

⁹ Nazwa „biały szum” odwołuje się do widma świetlnego, które stanowi mieszaninę widm postrzeganych przez człowieka jako kolorowe. Jest to w istocie zakłócenie losowe o rozkładzie normalnym (por. *Aneks*).

¹⁰ Prowadzone w ostatnich latach badania sekwencyjnych struktur w tekstach łacińskich, greckich, polskich i angielskich wskazują, że generowane z tekstów szeregi są zawsze stacjonarne (por. Pawłowski 1998, 2001, 2003; Pawłowski, Eder 2001; Pawłowski et al. 2010).

ufności dla takiego rozkładu. Używając statystycznej terminologii, intencją Woroncza było potraktowanie tekstu jako generatora zmiennej losowej, powiązanej ze strukturą kolejnych wersów. Wartością tej zmiennej miały być długości wersów, wyrażone liczbą sylab. Jako serię odpowiadającą symbolicznemu zapisowi AAA... traktował on przystające do siebie wersy tej samej długości, podczas gdy sekwencje wersów o różnych długościach traktowane były jako odpowiadające symbolicznemu zapisowi BBB... Woroncza mówił w tym wypadku o asymilacyjnej (AAA...) lub dysymilacyjnej (BBB...) tendencji długości przystających wersów. Przykład kodowania tekstu *Kroniki Dalimila* tą metodą (pierwsze dwanaście wierszy rozdziału 6) przedstawia tabelka 1.

Když sobě ten div ukázachu,	9	B
na Přemyslu potázachu,	8	A
keré by bylo znamenie	8	
té suché otky vzektivnie.	7	B
Jim tak Přemysl odpověď,	8	A
řka: „To jáz vám všě pověď.	8	
Otká suchá jest' znamenie	8	
mého chłapieho urozenie.	9	B
Ale že-t' jest' brzo vzkvetła,	7	
jakž vem Liubušě byla řekła,	9	
mój rod z chłapieho pořáda	8	A
dojde králového řáda;	8	

Tab. 1. Kodowanie tekstu *Kroniki Dalimila* według metody Woroncza.

Pewna słabość tego podejścia polega na tym, że zakłada się, iż tekst wersyfikowany – powstający zawsze jako efekt świadomej aktywności intelektualnej człowieka i wzbogacony o naddane (a więc także porządkujące) elementy estetyczne – porównywany jest z produktem abstrakcyjnie pojmowanego generatora liczb losowych. O ile jednak ten zarzut może się wydać dyskusyjny (w istocie odniesienia do struktur losowych są w statystyce lingwistycznej stosowane), inna kwestia jest poważniejsza. Natura testu serii sprawia, że teksty różne, zawierające jednak identyczną lub podobną liczbę różnie uporządkowanych serii AAA... i BBB..., dadzą praktycznie takie same parametry N , n_a i n_b . Z tego względu test serii nie został użyty w naszych badaniach, których rezultaty zostały przedstawione i przedyskutowane poniżej.

6. Przygotowanie i kwantyfikacja danych

Badania ilościowe języka prowadzi się na ogół metodą reprezentacyjną. Jednak w przypadku tekstów krótkich, kiedy kodowanie nie jest nadmiernym obciążeniem, może być również stosowana indukcja zupełna. Tak właśnie postąpiono analizując tekst *Kroniki Dalimila*. Z uwagi na to, że w użytej przez nas edycji *Kronika* składa się ze stu sześciu rozdziałów poprzedzonych jednym rozdziałem wstępnym, uznano ten podział za znaczący pod względem stylometrycznym i przyjęto, że tekst może być reprezentowany szeregiem czasowym stu sześciu wartości, z których każda odpowiada średniej długości wersu w rozdziale, mierzonej liczbą sylab (w celach porównawczych do pomiaru średniej długości wersu użyto także liczby liter i wyrazów).

Analizując układy rytmiczne, tworzone przez kolejne sylaby tekstu, zastosowano natomiast metodę reprezentacyjną. W obrębie każdego rozdziału wybierano losowo fragmenty długości ok. stu pięćdziesięciu sylab, które były kodowane jako sekwencje akcentów dynamicznych lub też iloczasów. W obu wypadkach stosowano więc skalę binarną, przypisując odpowiednio wartość 1 sylabom akcentowanym i 0 nieakcentowanym. Podobnie postąpiono z kodowaniem sekwencji sylab długich (1) i krótkich (0). Warto w tym miejscu zwrócić uwagę na problematyczność kodowania na potrzeby badań współczesnych tekstu pochodzącego z epoki średniowieczna. Jakikolwiek błędy w tej materii mogłyby zdeformować wynik, nawet przy zastosowaniu do obliczeń wymyślnych technik matematycznych. Chociaż po upływie siedmiuset lat dokładne odtworzenie brzmienia *Kroniki Dalimila* jest raczej niewykonalne, a z pewnością nieweryfikowalne, sekwencję samych akcentów dynamicznych i iloczasów można zrekonstruować, znając ogólne zasady prozodii i morfologii konkretnego języka. Struktury prozodyjne w języku opierają się bowiem na układzie sylab, a więc jednostek poddających się stosunkowo łatwej i przewidywalnej segmentacji.

7. Rezultat badań i interpretacja

Testowanie hipotezy prozaizacji *Kroniki Dalimila* rozpoczęto od obliczenia kilku podstawowych miar statystycznych (średniej, wariancji i odchylenia standardowego) w kolejnych rozdziałach tekstu. Celem była wstępna eksploracja danych, ukierunkowana na rozpoznanie ich charakterystyki statystycznej. Wskazówki dotyczące kierunku ewolucji stylu *Kroniki*, jakie udało się wyprowadzić dzięki tej analizie, zostały potwierdzone późniejszymi badaniami.

Analizę szczegółową struktury wersyfikacyjnej rozpoczęto od najbardziej intuicyjnej miary, a mianowicie badania stabilności długości następujących po sobie

wersów. Nawet w tych rozdziałach, które zawierają wiersz nieregularny (cz. *bez-rozměrný verš*), należałoby spodziewać się względnej stabilizacji tego parametru – przede wszystkim w obrębie rozdziałów, ale także w kolejnych rozdziałach. Wykres pierwszy (Rys. 1) wskazuje, że średnia długość wersu rośnie od początku do końca *Kroniki Dalimila* od ośmiu do około dwunastu sylab, przy czym w części środkowej widać wyraźne zaburzenie tej tendencji¹¹. Taką samą zależność można zaobserwować na wykresie drugim (Rys. 2), gdzie zastosowano jako miarę długości wersu wyrazy (podobny wynik dały także długości mierzone w literach). Zależności te można wyrazić następującym modelem regresji liniowej:

$$(1) \quad y_i = x_i \beta_1 + \beta_0 + \varepsilon$$

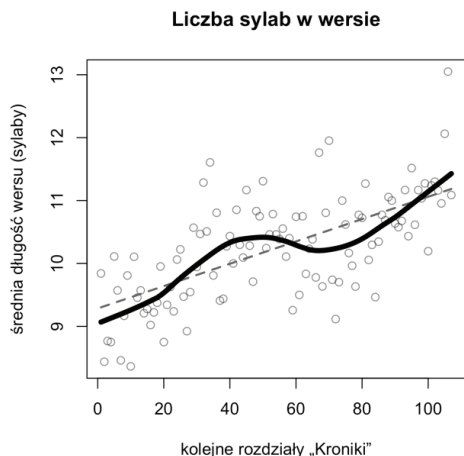
gdzie β_1 β_0 – współczynniki modelu,
 ε – element losowy (szum).

Po estymacji parametrów β_1 i β_0 otrzymano model wyrażony równaniem (2), gdzie x_i oznacza średnią długość wersu w i -tym rozdziale, a y_i jest numerem rozdziału. Pozytywna, chociaż niewielka, wartość współczynnika β_1 wskazuje na łagodny trend wznoszący średniej długości wersu (od dziewięciu sylab w początkowych rozdziałach do jedenastu w części końcowej).

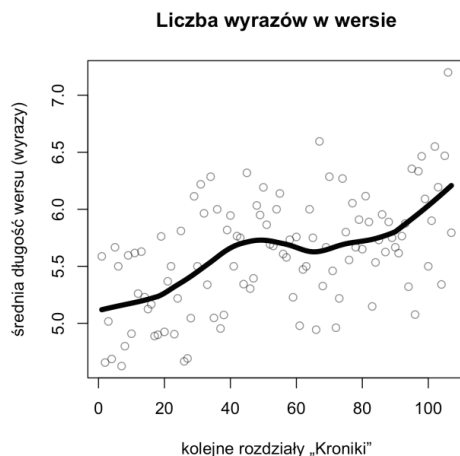
$$(2) \quad \hat{y}_i = 0.0177x_i + 9.2847 + \varepsilon$$

Model oparty na funkcji liniowej, zaznaczony na Rys. 1 linią przerywaną, jest statystycznie zadowalający, ale nie oddaje wyraźnie widocznego w połowie tekstu zachwiania tendencji. Charakter tego zaburzenia sugeruje, że jego źródłem nie była wewnętrzna dynamika procesu pisania, lecz jakiś zewnętrzny czynnik zakłócający stabilny rozwój narracji kronikarskiej.

¹¹ Linia trendu została wygenerowana z wykorzystaniem procedury opisanej przez W.S. Clevelanda (1979).



Rys. 1. Średnia długość wersów kolejnych rozdziałach *Kroniki* (liczba sylab w wersie).

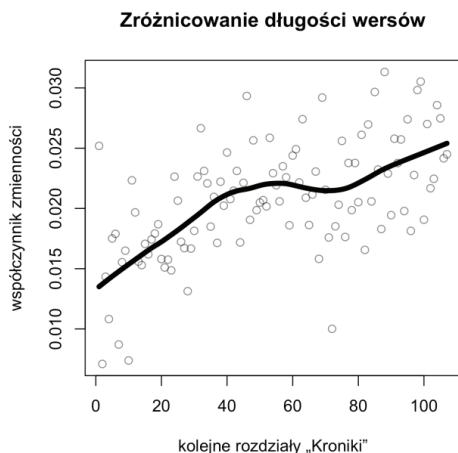


Rys. 2. Średnia długość wersów kolejnych rozdziałach *Kroniki* (liczba wyrazów w wersie).

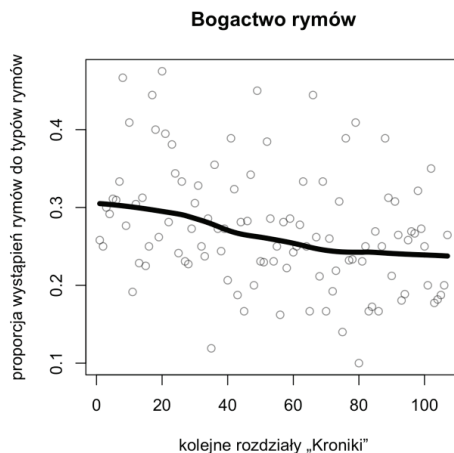
Powyższe wyniki same w sobie nie dają żadnej podstawy do wnioskowania o zmianie stylistycznej, polegającej na przekształcaniu się wiersza w prozę rymowaną. Jeśli jednak uzupełni się je innymi wskaźnikami, takimi jak miara rozrzutu długości wersu w kolejnych rozdziałach, ogólny obraz stanie się bardziej klarowny. Podstawową miarą rozrzutu jest odchylenie standardowe, tutaj jednak warto użyć jego wartości znormalizowanej, zwanej współczynnikiem zmienności (odchylenie standardowe podzielone przez średnią arytmetyczną). Jak widać na Rys. 3, również ta miara pokazuje dezintegrację długości wersów, które stają się coraz dłuższe w kolejnych rozdziałach *Kroniki*, a jednocześnie coraz bardziej chaotyczne (nieprzewidywalne) pod względem oczekiwanej długości wersów. I pod tym względem zaczynają przypominać bardziej prozę niż mowę wiązaną. Wyniki uzyskane dla sylab zostały potwierdzone w innych testach, przeprowadzonych dla długości wersów mierzonych w literach oraz w wyrazach. Jak się okazuje, jednostki segmentacji tekstu nie zmieniają ogólnego obrazu: zawsze występuje trend wzrostowy i zwiększająca się dyspersja.

Warto tutaj zauważyć, że kształt wersyfikacyjny kilku początkowych rozdziałów, w których dominuje ósmiozłogostkowiec z nielicznymi wystąpieniami nieco krótszych lub nieco dłuższych wersów, wpisuje się w bardzo bogatą i ciekawą tradycję ludowej poezji ustnej wielu kultur europejskich. W literaturze polskiej ten typ wiersza spotykamy na przykład w średniowiecznej *Legendzie o św. Aleksym* albo w *Rozmowie mistrza Polikarpa ze śmiercią*, później zaś w zachowanych XIX-wiecznych balladach ludowych. *Kronika Dalimila* wygląda więc tak, jakby autor układał jej najwcześniejsze fragmenty, korzystając z ustnej poezji epickiej,

po których następowało coraz więcej materiału przejmowanego z pozaliterackich (głównie historycznych) źródeł pisanych. Jeśli to przypuszczenie jest słuszne, powinno je potwierdzić zachowanie innych cech prozodyjnych *Kroniki*, takich jak rymy.



Rys. 3. Współczynnik zmienności średniej długości wersu w poszczególnych rozdziałach *Kroniki* (w sylabach).



Rys. 4. Bogactwo rymów mierzone jako proporcja wystąpień par rymowych to typów par rymowych w poszczególnych rozdziałach *Kroniki*.

Kolejne zatem testy miały za zadanie odpowiedzieć na szereg pytań związanych z homogenicznością rymów oraz z ich repertuarem (bogactwem). Z uwagi na fakt, iż archaiczne, ustne poematy epickie wykazują tendencję do wyrównywania rymów (czy też do redukcji częstych alternacji), co przekłada się na stosunkowo niewielki zasób końcówek i dłuższe ciągi wersów rymowanych identycznie – w przeciwieństwie do wysoce zróżnicowanych i rymowo bogatej poezji pisanej – szczegółowej analizie poddano następstwa rymów w *Kronice*.

Pierwszy test polegał na obliczeniu stosunku unikalnych typów par rymowych do wszystkich wystąpień takich par w danym rozdziale. Uzyskana tym sposobem wartość jest odpowiednikiem podstawowej miary bogactwa leksykalnego, czyli proporcji liczby typów do liczby okazów/wystąpień (ang. *type/token ratio*). Uzyskane wyniki wskazują na duży rozrzut bogactwa rymów w poszczególnych rozdziałach *Kroniki* (Rys. 4), jednak obecność rosnącego trendu jest tutaj dość wyraźnie widoczna. Otrzymany rezultat nie jest zgodny z intuicją, a przede wszystkim przeczy poczynionym powyżej obserwacjom, sugeruje bowiem, że większe bogactwo rymów występuje w początkowych rozdziałach *Kroniki*, a więc tych, któ-

re powinny wykazywać więcej cech literatury ustnej. Trzeba jednak pamiętać, że wartość *type/token ratio* jest czuła na długość badanej próby tekstowej i właśnie ta zmienna mogła wpłynąć na otrzymany rezultat.

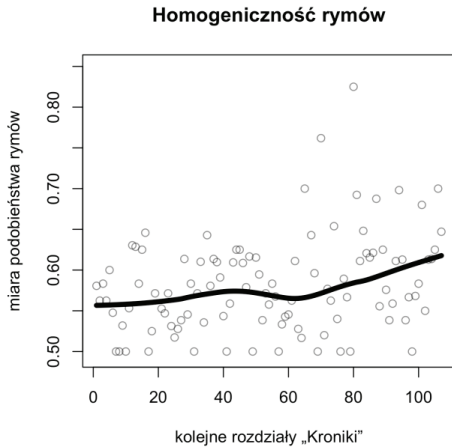
Następny test polegał na obliczeniu stosunkowo prostej miary homogeniczności par rymowych. W zamierzeniu miał on wychwycić pojawiające się czasem w *Kronice* dłuższe fragmenty, oparte na jednym i tym samym rymie. Sytuacje te widać oczywiście gołym okiem: czasem jeden rym obejmuje kilka (albo i kilkanaście) par wersów, po czym następuje zmiana i znowu dopiero po kilku następnych parach następuje kolejna alternacja. Wspomniana miara miała na celu porównanie stopnia homogeniczności rymów w poszczególnych rozdziałach. Jest ona skonstruowana w ten sposób, że liczba wersów powtarzających rym poprzedni bez alternacji podzielona jest przez liczbę wszystkich wersów występujących w danym rozdziale. Bezwzględna alternacja każdej kolejnej pary rymowej da zatem wartość 0,5, podczas gdy rozdział liczący n wersów oparty na jednym tylko rymie (brak alternacji) da w wyniku wartość homogeniczności równą $(n-1)/n$. Oczywiście obie powyższe miary, tj. indeks bogactwa rymów oraz indeks homogeniczności, są w pewnym stopniu skorelowane, ale pokazują dwa różne aspekty alternacji rymów. Na przykładzie poniższego fragmentu z rozdziału 38. łatwo wyjaśnić różnicę między obiema miarami:

Tehdy sě sta, že kněz Oľdřich łovieše
 a sám v pustém lesě blůdieše.
 Když u velikých túhách bieše,
 okoľo sebe všady zřieše,
 uzřě, nali-ť stojí dospěť hrad.
 Kněz k němu jíti chtieše rád.
 Ale že cěsty neumějieše
 a okoľo hľožíe husté bieše,
 ssěd s koně, mečem cěstu proklesti,
 i počě po ostrvách v hrad lězti;
 neb sě nemože nikohěho dovoľati,
 by v něm liudie byli, nemože znamenati.
 Most u něho vzveden bieše,
 a hrad zdi okoľo sebe tvrdé jmějieše.
 Když kněz s úsilím v hrad vnide
 a všecky sklepy znide,
 zetlelé rúcho vidieše,
 však i čľověka na něm nebieše.
 Sbožíe veliké a vína mnoho naleze.
 Ohledav hrad, kudyž byľ všěť, tudyž vyleze.
 Pak kněz hrad ten da pánu tomu, jemuž Přiemu diechu;
 proto tomu hradu Přimdu vzděchu.

W powyższym fragmencie znajduje się 12 wersów, w których rym jest taki sam, jak w wersie poprzednim, liczba jednak typów par rymowych wynosi 7, jako że rym *-ieše* powraca wiele razy. Tym samym obie miary będą się różniły w sposób znaczący: miara bogactwa rymów wyniesie 0,318, a miara homogeniczności rymów 0,545.

Na Rys. 5 pokazana została miara homogeniczności rymów dla kolejnych rozdziałów *Kroniki*. Zarówno pewien rozrzut wyników, jak i obecność kilku obserwacji nietypowych (rozdziałów ponadprzeciętnie homogenicznych pod względem rymów) nie zmieniają faktu, że bez większych przeszkód daje się zauważyć ogólny trend wzrostowy. Początkowe rozdziały charakteryzuje nie tylko większa różnorodność używanych par rymowych w porównaniu z rozdziałami końcowymi, ale także unikanie łączenia homonimicznych rymów w dłuższe ciągi. Można by w pewnym uproszczeniu powiedzieć, że wyszukane zróżnicowanie rymów i ich konsekwentna alternacja (a więc cechy, które tradycyjnie wiąże się z wysokoartystyczną poezją epoki piśmiennej) powoli zmieniają się w powtarzalność, przewidywalność, pewnego rodzaju formuliczną „jałowość”, tak charakterystyczną dla archaicznej twórczości ustnej. Kolejny bardzo prosty test potwierdza powyższe uwagi: jeśli obliczyć względne frekwencje wszystkich typów rymów (a więc dla każdego typu podzielić liczbę jego wystąpień przez liczbę wersów w danym rozdziale) i poddać bliższemu oglądowi np. 5 najczęstszych typów, ujawni się bardzo ciekawa tendencja (Rys. 6). Widać mianowicie dość wyraźnie, że kilka najczęstszych par rymowych stopniowo „wypiera” z pozycji rymowej – z każdym kolejnym rozdziałem *Kroniki* – różne inne typy. Pozycja rymowa pod koniec tekstu staje się więc coraz bardziej przewidywalna; co więcej, oznacza to również stopniowe upraszczanie i stypizowanie składni w końcowych partiach *Kroniki*, jako że owe pleniące się przewidywalne rymy to w przeważającej większości zaledwie kilka morfemów, w tym końcówki fleksyjne czasowników (*-echul/-ichul/-achu*, *-eše/-aše*, *-atil/-eti*, *-idul/-edul/-adu*) oraz dopełniacz liczby mnogiej rodzaju męskiego rzeczowników, przymiotników i zaimków (*-éhol/-oho*). Ów zdumiewająco mały zestaw końcówek zajmuje w końcowych rozdziałach nawet 60% wszystkich pozycji rymowych, świadcząc w sposób oczywisty o redukcji składni do kilku zaledwie powtarzających się konstrukcji.

Wymowę zaobserwowanego tutaj zjawiska osłabia jednak wspomniana wyżej zależność wartości współczynników bogactwa i homogeniczności od długości wersów. W miarę stabilny zestaw rymów w próbach o coraz większej długości będzie niejako automatycznie zwiększał homogeniczność tekstu (rozumianą zgodnie z podaną wyżej definicją).



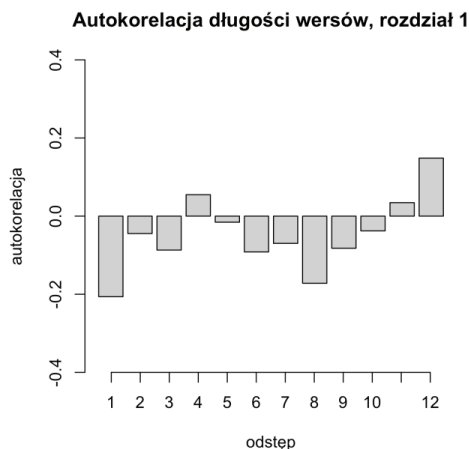
Rys. 5. Skupianie się homogenicznych par rymowych w dłuższe ciągi w kolejnych rozdziałach *Kroniki*.



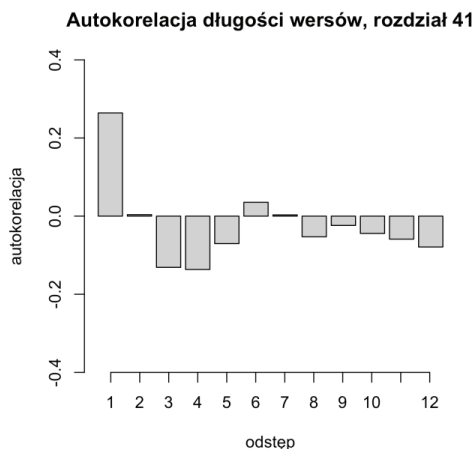
Rys. 6. Częstość względna pięciu najczęstszych par rymowych w kolejnych rozdziałach *Kroniki*.

Nadspodziewanie wyraźne trendy, widoczne przy zastosowaniu prostych miar statystycznych, zachęcają do użycia technik analizy sekwencyjnej, pozwalających na dostrzeżenie wewnętrznych korelacji pomiędzy następującymi po sobie wersami *Kroniki*. Celem poniższych testów było przede wszystkim sprawdzenie, czy rzeczywiście krótkie wersy mają tendencję do następowania po długich (i odwrotnie), a zatem czy następuje zauważalna alternacja długich i krótkich (Hrabák 1959). Jako hipotezy alternatywne przyjęto brak znaczącej autokorelacji długości wersów (w takim wypadku nie istnieje żadna zasada następstwa wersów ze względu na długość) lub istnienie autokorelacji pozytywnej, oznaczającej częste następstwo wersów tej samej długości. W celu zweryfikowania powyższych twierdzeń (można je określić mianem hipotezy Hrabáka) dla poszczególnych rozdziałów została obliczona funkcja autokorelacji (ACF) dla długości wersów wyrażonych w sylabach. Następnie oszacowano wartość funkcji autokorelacji cząstkowej (PACF), co pozwoliłoby na znalezienie odpowiednich parametrów modelu autoregresji (ewentualnie modelu ruchomej średniej) oraz na sprawdzenie stopnia dopasowania modelu. Opisana powyżej procedura została przeprowadzona niezależnie dla każdego rozdziału *Kroniki*¹². Chociaż ten etap analizy szeregów czasowych jest najważniejszy, nie zawsze dało się go zastosować do danych z *Kroniki Dalimila*, jako że w wielu wypadkach wyniki nie wykazywały żadnych istotnych statystycznie korelacji.

¹² Ponieważ modelowanie przy użyciu techniki ARIMA jest bardzo czasochłonne, zdecydowano się na analizę co drugiego rozdziału *Kroniki*. Wolno jednak wierzyć, że wyniki uzyskane przy takim podejściu dają dobre przybliżenie do tego, co można uzyskać przez analizę całego zbioru danych.

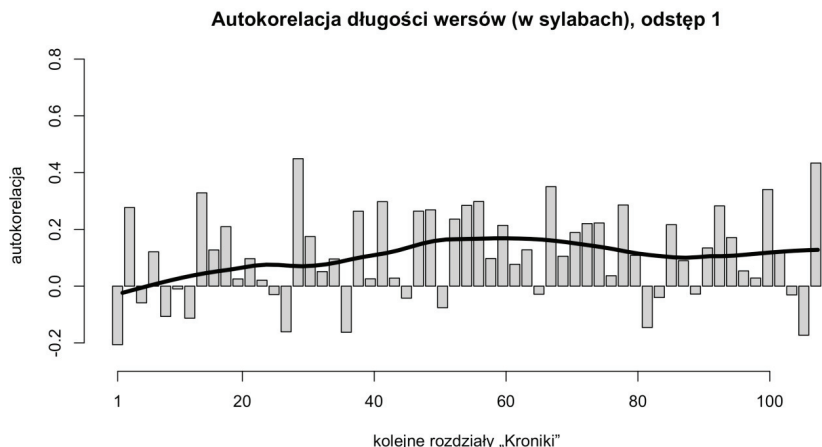


Rys. 7. Funkcja autokorelacji (ACF) sekwencji długości wersów w rozdziale 1.



Rys. 8. Funkcja autokorelacji (ACF) sekwencji długości wersów w rozdziale 41.

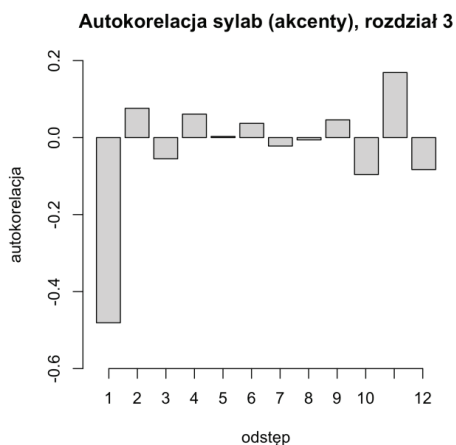
Uzyskane wyniki okazały się ciekawe, ale poziom ich istotności statystycznej był niski. Jak widać na Rys. 7 (wyniki dla rozdziału 1), funkcja autokorelacji nie ujawnia żadnych znaczących wartości. Jedynie bardzo bliska negatywna korelacja o odstępie 1 (dotycząca dwóch sąsiadujących wersów) warta jest bliższego oglądu. Owa negatywna wartość oznacza, że dowolny wers jest z dużym prawdopodobieństwem odmienny od wersu poprzedzającego: krótki wers najczęściej będzie wymuszał następujący po nim wers długi, a po długim najczęściej wystąpi wers krótki. Brak korelacji długookresowych daje się wyłożyć lingwistycznie w ten sposób, że dowolny wers jest w jakimś stopniu zależy od poprzedniego, ale owa „pamięć” nie sięga głębiej – długość wersu nie wynika z dłuższego kontekstu metrycznego. Funkcja ACF oszacowana dla następnych rozdziałów potwierdza ogólnie słabe wyniki uzyskane dla rozdziału 1. Co więcej, jedyna znacząca wartość dla odstepu 1 jest dla niektórych rozdziałów negatywna (Rys. 7), dla innych zaś pozytywna (Rys. 8), z dużą tendencją do zmienności. Z tego powodu zastosowano nietypową procedurę, a mianowicie zestawiono na jednym wykresie wszystkie wartości ACF dla odstepu 1 w kolejnych rozdziałach. Pozwoliłoby to stwierdzić istnienie ewentualnych trendów lub fakt taki wykluczyć. Okazało się, że mamy tu do czynienia z powolnym trendem rosnącym wartości autokorelacji (Rys. 9): wersy w rozdziałach końcowych są nieco bardziej rytmiczne niż w początkowych, mimo że – paradoksalnie – ta tendencja idzie zupełnie w przeciwnym kierunku niż opisana powyżej tendencja do prozaizacji. Innymi słowy: rozdziały początkowe są metrycznie bardziej uporządkowane, a końcowe upodobią się bardziej do prozy, mimo że ich rytmiczność, mierzona długościami wersów, rośnie.



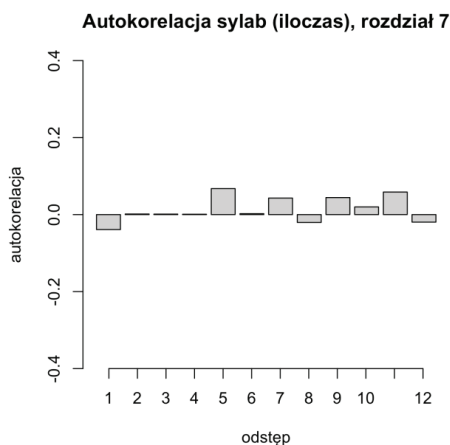
Rys. 9. Wybrane wyniki funkcji autokorelacji (wyłącznie odstęp 1) dla poszczególnych rozdziałów *Kroniki*.

Mówiąc o rytmiczności, nie wolno pominąć najbardziej oczywistego typu danych, które doskonale nadają się do modelowania metodami analizy sekwencyjnej. Mowa o wierszu zakodowanym jako sekwencja nacechowanych i nienacechowanych sylab (lub innych jednostek prozodyjnych, np. stóp metrycznych). W zależności od typu wiersza tymi nacechowanymi i nienacechowanymi jednostkami będą sylaby akcentowane i nieakcentowane bądź też sylaby długie i krótkie. W wypadku *Kroniki Dalimila* – i języka czeskiego w ogólności – za prozodję odpowiadają w zasadzie obie cechy. Takie ukształtowanie czeskiego wiersza daje rzadką możliwość porównania obu tych systemów prozodyjnych ze sobą.

W celu przeprowadzenia obliczeń dane tekstowe zostały zakodowane jako binarne ciągi nacechowanych i nienacechowanych jednostek na dwa sposoby: raz jako szeregi czasowe akcentowe, raz jako szeregi iloczynowe. Wyniki działania funkcji ACF dla przykładowego rozdziału *Kroniki* zostały przedstawione na Rys. 10 (akcent) oraz Rys. 11 (iloczas). Oba wykresy są bardzo proste w interpretacji. Szereg akcentowy zawiera jedną silną negatywną korelację o odstepie 1 i brak korelacji dalszego stopnia, co oznacza, że dowolna sylaba z dużym prawdopodobieństwem przyjmie przeciwną wartość cechy w stosunku do sylaby poprzedzającej. Zjawisko to w sposób oczywisty wynika z dominującej alternacji sylab akcentowanych i nieakcentowanych w badanej próbce – nic dziwnego, że podobny wzorec odnaleziono we wszystkich pozostałych próbkach, a w kolejnych rozdziałach *Kroniki* nie występuje żadna ilościowa zmiana tej cechy. Wydaje się więc, że mamy tu do czynienia z cechą systemową, typową dla języka, a nie dla konkretnego tekstu. Nie wnosi ona niczego do stylistycznej analizy *Kroniki*, nie niesie też żadnej dodatkowej informacji.



Rys. 10. Funkcja autokorelacji (ACF) dla sekwencji sylab akcentowanych dynamicznie (rozdział 3).



Rys. 11. Funkcja autokorelacji (ACF) dla sekwencji sylab długich i krótkich (rozdział 7).

Szereg czasowy kodowany według iloczasu sylab daje wyniki skrajnie odmienne od omówionego powyżej szeregu akcentowego. Mimo że istnienie iloczasu w starożytnym jest faktem bezspornym, niniejsza analiza pokazuje, że jest on nieistotny jako relewantna cecha prozodyjna poezji, budująca jej rytmikę. Brak jakichkolwiek korelacji na Rys. 11 (pomijając wartości nieistotne, będące w istocie szumem) jest ewidentny i jednoznaczny. Obie cechy prozodyjne, które na ogół kojarzone są z wierszem, wersyfikacją i technikami poetyckimi – a więc iloczasy i akcent – okazały się cechami systemowymi zupełnie oderwanymi od wyborów autorskich: nie odegrały żadnej roli w stopniowej zmianie stylistycznej, zaobserwowanej w tekście *Kroniki Dalimila*.

8. Wnioski

Analizę struktury wersyfikacyjnej *Kroniki Dalimila*, przeprowadzona ilościowymi metodami stylometrii, potwierdziła częściowo hipotezę Jerzego Woronczaaka, wyrażoną w 1963 roku, mówiącą o postępującej prozaizacji tekstu. Uzyskane wyniki badań pozwoliły odpowiedzieć również na nowe pytania badawcze, dotyczące tekstu *Kroniki*. Jedno z nich miało związek z rolą kultury oralnej w tworzącym się nowym uniwersum piśmienniczym wieków średnich. Chociaż w pracy Woronczaaka nie ma bezpośrednich odwołań do prac J. Lorda, M. Parry’ego, W. Onga czy E. Havelocka, a więc najczęściej cytowanych badaczy oralności¹³,

¹³ Wyrażenie „literatura oralna”, spotykane często w tekstach naukowych, jest oksymoronem, ponieważ logiczną presupozycją istnienia literatury jest piśmienność (łac. *litteratura* to przecież

rozważania tego wrocławskiego badacza literatury średniowiecza są całkowicie spójne z ich poglądami. Woronczak przyjmuje za naturalne, że w „zapisanej literaturze oralnej” występować będzie tendencja do utrzymywania tej samej długości i prostej struktury powtarzanych segmentów rytmicznych (w zapisie odpowiadają im wersy), a więc zjawiska typowe dla oralności¹⁴.

Warto zastanowić się, co mogło być przyczyną stosowania tej monotonnej wersyfikacji, która w przypadku *Kroniki Dalimila*, ale także innych utworów, powstałych w fazie przechodzenia kultury europejskiej od oralności do piśmienności, pojawia się w niektórych fragmentach *Kroniki* i tutaj została przyjęta za swoisty punkt odniesienia dla badań empirycznych i wnioskowań. Z oczywistych względów trudno udzielić na to pytanie odpowiedzi uzasadnionej twardym materiałem dowodowym. Jednak najprawdopodobniej chodziło o to, że wykonawcom (bardom, guślarzom) naturalny i monotony rytm ułatwiał, jako czynnik mnemotechniczny, zapamiętanie i odtwarzanie z pamięci setek lub tysięcy wersów. Natomiast publiczność wsłuchana w epickie dzieje pradawnych królów i herosów chętnie poddawała się hipnotyzującemu działaniu rytmicznej melorecytacji, wspartemu dodatkowo dźwiękami towarzyszącego instrumentu (lutni, bębenka). Gwałtowne zmiany czy załamania powtarzalnej i przewidywalnej struktury rytmicznej występowały jedynie w krytycznych momentach fabuły.

Jednak uproszczający wniosek, że styl i kompozycję *Kroniki* ukształtowała opozycja oralności i piśmienności, nie wyjaśnia w dostatecznym stopniu występujących w środkowych partiach zaburzeń stabilnej tendencji wzrostu długości wersów (por. Rys. 1 i 2). Należy więc powyższą interpretację rozszerzyć, wzbogacając ją o nowe elementy. Otóż najprawdopodobniej pierwsze partie *Kroniki* powstawały jako kompilacje lub dosłowne cytaty poezji oralnej, istniejącej długo przed powstaniem wersji pisanej. Świadczyłyby o tym użycie archaicznego metrum ośmiozłotkowego, które jest typowe dla poezji oralnej. Autor *Kroniki* starał się zapewne, w miarę jak opisywane zdarzenia stawały się coraz bardziej współczesne, pozostać wierny temu metrum po to, by zagwarantować formalną i estetyczną spójność swojego dzieła. Jednak konieczność coraz dokładniejszego, wierniejszego i – w pewnym sensie – epickiego opisu, okazała się wyzwaniem nie do pokonania. Innymi słowy – połączenie kompetencji historyka i talentu poetyckiego – nader rzadkie w ciągu wieków – najwidoczniej nie było atrybutem także anonimowego *Dalimila*. Załamanie linii trendu w środkowej części *Kro-*

pismo), a więc przeciwieństwo oralności. Wyrażenie to można więc przyjąć jedynie jako ogólnie akceptowaną metaforę opisywanego zjawiska.

¹⁴ Potwierdzają to wrażenia audytywne, jakie wywiera odsłuchiwanie słynnego zbioru nagrań guślarzy, dokonanych przez J. Lorda i M. Parry'ego w latach 30. XX wieku na obszarze Jugosławii (por. <http://chs119.chs.harvard.edu/mpc/>).

niki świadczyłyby więc o nieudanych próbach ujęcia nowych, bogatszych treści, w starym metrum. Być może autor w końcu uświadomił sobie, że zadanie to przekracza jego siły, a i przekazywane treści nie odpowiadają już uproszczonej formie, stosowanej przez poprzedników: takim sposobem poeta opowiadający bajeczne dzieje królestwa Czech przeistoczył się w historiografa, starającego się pisać wierszem, a w istocie piszącego prozą rymowaną.

Nie oznacza to, że inne interpretacje uzyskanych wyników można wykluczyć. Istnieje przecież pewne prawdopodobieństwo, że *Kronika* miała wielu autorów i różne jej partie mają z tego powodu inne metrum. Zmiany stylistyczne w jej środkowych partiach byłyby więc efektem indywidualnych upodobań piszących. Jednak istnienie trendu, wprawdzie zakłóconego, lecz widocznego w całości *Kroniki*, wydaje się podważać taką interpretację.

Za poznawczo inspirujące należy też uznać spostrzeżenia, dotyczące niewielkiej roli iloczasu w budowaniu rytmiki tekstu staroczeskiego. Wynik ten jest spójny z obserwacjami, jakie udało się poczynić we wcześniejszych badaniach na materiale łacińskim, różni się jednak od rezultatów otrzymanych z badań klasycznego heksametru greckiego. Kwestia jest jednak złożona, ponieważ analiza tekstu pisanego, a nie jego autentycznego odczytania, pomija czas trwania sylab i/lub samogłosek. Z prac na temat prozodii wiadomo, że można wyróżnić języki o izochronii akcentowej (ang. *stress-timed*), izosylabiczne (ang. *syllable-timed*) oraz mieszane (por. Dauer 1983, Bertinetto 1989, Abercrombie 1967). Podział ten opiera się na ogólnej hipotezie tzw. izochronii językowej, zakładającej istnienie rytmu mowy, opartego na regularnej powtarzalności segmentów o jednakowym czasie trwania. Języki z izochronią akcentową charakteryzowałyby się tendencją do utrzymywania stałych odstępów międzyakcentowych (w zależności od języka jest to kilkaset milisekund), powodującą w konsekwencji skracanie lub wydłużanie pojedynczych głosek. Natomiast w językach izosylabicznych istniałaby tendencja do zachowywania względnie stałych długości sylab. Trudno dziś dociekać, jaka była charakterystyka staroczeskiego (czeszczyzna nowożytna jest uznawana za język o izochronii mieszanej, por. Dankovičová, Dellwo 2007). Można jednak powiedzieć z całą pewnością, że wiarygodne badanie mogłoby dotyczyć wyłącznie oryginalnego wykonania, a to skazuje współczesnego badacza na zbyt daleko idące domysły. W tej sytuacji symulacja takiego odczytania, oparta na zapisie filologów badających wnikliwie teksty źródłowe, wydaje się podejściem uzasadnionym metodologicznie, a na dodatek jedynym możliwym.

Podsumowując, można powiedzieć, że wyniki przeprowadzonych badań statystycznych tekstu *Kroniki Dalimila* dostarczyły nowej wiedzy na temat jej wewnętrznej struktury i metryki. Natomiast interpretacje dotyczące współzależności oralności i piśmienności oraz prawdy faktograficznej i fikcji wykraczają poza średniowieczne uniwersum i potwierdzają istnienie bardzo ogólnych, niezmiennych zasad konstruowania narracji opisowej.

Teksty źródłowe

- Jireček J., (red.), 1882, *Rýmovaná kronika česká tak řečeného Dalimila*. Vyd. J. Jireček, [w:] *Prameny dějin českých*, vol. 3, part 1, Praha.
- Havránek B., Daňhelka J. (red.), 1957, *Nejstarší česká rýmovaná kronika tak řečeného Dalimila*, Praha: Nakladatelství Československé akademie věd.
- Daňhelka et al, Daňhelka J., Hádek K., Havránek B., Kvítková N., (red.), 1988, *Staročeská kronika tak řečeného Dalimila*. Vydání textu a veškerého textového materiálu, vol. 1–2. Praha: Academia.

Literatura

- Abercrombie D., 1967, *Elements of General Phonetics*, Edinburgh University Press.
- Bertinetto P.M., 1989, *Reflections on the Dichotomy ‘Stress’ vs. ‘Syllable-timing’*, „Revue de Phonétique Appliquée”, vol. 91–93, pp. 99–130.
- Box G., Jenkins G., 1976, *Time series analysis: forecasting and control*, San Francisco: Holden-Day.
- Cleveland W.S., 1979, *Robust locally weighted regression and smoothing statterplots*, „Journal of the American Statistical Association”, vol. 74, pp. 829–836.
- Cryer J., 1986, *Time Series Analysis*, Boston: Duxbury Press.
- Culley R.C., 1986, *Oral Tradition and Biblical Studies*, „Oral Tradition” 1/1, pp. 30–65.
- Dankovičová J., Dellwo V., 2007, *Czech Speech Rhythm and the Rhythm Class Hypothesis*, [In:] *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, pp. 1241–1244. [<http://www.icphs2007.de/conference/Papers/1538/1538.pdf> – dostęp 02.06.2015].
- Dauer R.M., 1983, *Stress-timing and Syllable-timing Reanalysed*, „Journal of Phonetics”, vol. 11, pp. 51–62.
- Eder M., 2008, *How rhythmical is hexameter: a statistical approach to Ancient epic poetry*, [In:] *Digital Humanities 2008: Book of Abstracts*, University of Oulu, pp. 112–114.
- Glass G.V., Wilson V.L., Gottman J.M., 1975, *Design and Analysis of Time-Series Experiments*, Colorado: Colorado Associated University Press.
- Gottman J.M., 1981, *Time-Series Analysis: A Comprehensive Introduction for Social Scientists*, Cambridge etc.: Cambridge University Press.
- Gottman J.M., 1990, *Sequential analysis*, Cambridge etc.: Cambridge University Press.
- Gregson R.A.M., 1983, *Time series in psychology*, Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Havelock E.A., 2006, *Muza uczy się pisać. Rozważania o oralności i piśmienności w kulturze Zachodu*, Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Herdan G., 1966, *Language as Choice and Chance*, Berlin etc.: Springer.
- Hrabák J., 1959, *Studie o českém verši*, Praha: SPN.

- Köhler R., 2005, *Synergetic linguistics*, [w:] *Quantitative Linguistik / Quantitative Linguistics. Ein Internationales Handbuch = An International Handbook*, eds. R. Köhler, G. Altmann, R. Piotrowski, Berlin, New York: Walter de Gruyter.
- Lutosławski W., 1897a, *The origin and growth of Plato's logic*, London : New York : Bombay: Longmans, Green and Co. [Reprint: Hildersheim, Georg Olms Verlag, 1983].
- Lutosławski W., 1897b, *Lutosławskis Theorie der Stylometrie auf die Platonische frage angewendet*, „Zeitschrift für Philosophie und philosophische Kritik”, Bd. 110, S. 171–219.
- Lutosławski W., 1898, *Principes de stylométrie appliqués à la Chronologie des œuvres de Platon*, „Révue des études grecques”, t. 41, pp. 61–81.
- McCleary R., Hay R.A., 1980, *Applied Time Series Analysis for the Social Sciences*, Beverly Hills etc.: Sage Publications.
- Pawłowski A., 1998, *Séries temporelles en linguistique. Avec application à l'attribution de textes: Romain Gary et Émile Ajar*, Paris and Genève: Champion-Slatkine.
- Pawłowski A., 2001, *Metody kwantytatywne w sekwencyjnej analizie tekstu*, Warszawa: Katedra Lingwistyki Formalnej Uniwersytetu Warszawskiego.
- Pawłowski A., 2003, *Sequential analysis of versified texts in fixed- and free-accent languages: example of Polish and Russian*, [In:] *Europa der Sprachen: Sprachkompetenz – Mehrsprachigkeit – Translation. Akten des 35. Linguistischen Kolloquiums in Innsbruck 2000. Teil 2: Sprache und Kognition*, ed. L.N. Zybatow, Frankfurt/M. etc.: Peter Lang, pp. 235–246.
- Pawłowski A., 2005, *Modelling of the sequential structures in text*, [w:] *Quantitative Linguistik / Quantitative Linguistics. Ein Internationales Handbuch = An International Handbook*, eds. R. Köhler, G. Altmann, R. Piotrowski, Berlin, New York: Walter de Gruyter, pp. 738–750.
- Pawłowski A., 2006, *Chronological analysis of textual data from the 'Wrocław Corpus of Polish'*, „Poznań Studies in Contemporary Linguistics”, vol. 41, pp. 9–29.
- Pawłowski A., Eder M., 2001, *Quantity or stress? Sequential analysis of Latin prosody*, „Journal of Quantitative Linguistics”, vol. 8, № 1, pp. 81–97.
- Pawłowski A., Eder M., 2015, *Sequential structures in "Dalimil's Chronicle"*, [In:] *Sequences in Language and Text*, eds. G.K. Mikros, J. Mačutek, Berlin, Boston: Walter de Gruyter, pp. 147–170.
- Pawłowski et al., 2010, Pawłowski A., Krajewski M., Eder M., *Time series modelling in the analysis of Homeric verse*, „Eos”, vol. 47, № 1, pp. 79–100.
- Woronczak J., 1993 [1963], *Zasada budowy wiersza „Kroniki Dalimila”*, [w:] *Studia o literaturze średniowiecza i renesansu*, red. J. Woronczak, Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, s. 67–75. [Pierwsze wydanie: „Pamiętnik Literacki”, t. 54 (1963), zes. 2, s. 469–478].

Summary

Sequential structures in “Dalimil’s Chronicle”: a stylometric analysis

The object of this paper is a quantitative study of sequential structures in the medieval Czech chronicle *Dalimilova Kronika*. The authors analyse style changes in the chronicle and try to answer some questions concerning its authorship. Another topic discussed in this paper concerns the relationship between orality and literacy at the threshold of the Middle Ages in Europe. A philological approach, combined with quantitative tools including trend analysis and time series modeling, is applied in this paper.

Keywords: Dalimil, Dalimil’s Chronicle, sequential analysis, stylometry, quantitative analysis.

Aneks

Szeregiem czasowym nazywamy sekwencję wartości x_i , reprezentujących dyskretną zmienną losową X_t , uporządkowanych na osi czasowej lub innej, ale mającej analogiczne właściwości (przykładem jest linearny porządek tekstu) X_t :

$$(1) \quad X_t = \{x_1, x_2 \dots x_i\}.$$

Średnia μ_x szeregu czasowego X_t zdefiniowana jest następująco:

$$(2) \quad \mu_x = E(X_t),$$

estymator średniej μ_x obliczany jest na podstawie wzoru:

$$(3) \quad m_x = \frac{1}{N} \sum_{t=1}^N x_t$$

gdzie N – długość szeregu

x_t – wartość szeregu w momencie lub na pozycji t .

Wariancja σ_x^2 szeregu czasowego X_t definiowana jest wzorem:

$$(4) \quad \sigma_x^2 = E(X_t - \mu_x)^2,$$

estymator wariancji σ_x^2 ma postać (oznaczenia jak wyżej):

$$(5) \quad s_x^2 = \frac{1}{N} \sum_{t=1}^N (x_t - m_x)^2.$$

Funkcja autokowariancji γ_k szeregu czasowego X_t definiowana jest jako:

$$(6) \quad \gamma_k = E\{(X_t - \mu_x)(X_{t+k} - \mu_x)\},$$

gdzie μ_x – średnia szeregu

k – odstęp oddzielający wartości x_t .

Autokowariancja szeregu czasowego estymowana jest wzorem:

$$(7) \quad c_k = \frac{1}{N-k} \sum_{t=1}^{N-k} (x_t - m_x)(x_{t+k} - m_x),$$

gdzie N – długość szeregu

x_t – wartość szeregu w momencie lub na pozycji t

k – odstęp oddzielający wartości x_t .

Funkcja autokorelacji (ACF) szeregu czasowego rzędu k definiowana jest jako:

$$(8) \quad \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_x^2},$$

gdzie γ_k – funkcja autokowariancji (jeżeli $k = 0$, wówczas $\gamma_k = \sigma_x^2$)
 k – odstęp oddzielający wartości x_t .

ACF estymowana jest następującym wzorem:

$$(9) \quad r_k = \frac{c_k}{c_0} = \frac{c_k}{s_x^2}.$$

Funkcja autokorelacji jest najważniejszym parametrem w analizie szeregu czasowego. Jeżeli jej wartość jest (według przyjętego kryterium) nielosowa, uzasadniona poznawczo jest estymacja procesu stochastycznego, jaki badana zmienna losowa realizuje. Jeżeli natomiast autokorelacja szeregu jest zerowa lub nieznacząca, przyjmuje się, że nie istnieje dająca się wyodrębnić składowa deterministyczna (innymi słowy w tekście lub w innych danych nie istnieje „pamięć” stanów poprzedzających).

Istnieją trzy podstawowe modele szeregów czasowych, a także wiele modeli złożonych. Ponieważ w tekście *Kroniki Dalimila* nie stwierdzono istnienia tych ostatnich, poniżej przedstawione zostaną tylko modele proste.

Proces losowy składa się ze statystycznie niezależnych realizacji zmiennej losowej $X_t = \{e_1, e_2, \dots\}$. Niezależność oznacza, że charakteryzuje go zerowa kowariancja i autokorelacja. Przez analogię do spektrum świetlnego wartości e_t o rozkładzie normalnym określane są także jako „biały szum” (ang. *white noise*).

Procesem autoregresji rzędu p nazywamy sekwencję wartości x_t zdefiniowanych wzorem:

$$(10) \quad x_t = a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_p x_{t-p} + e_t,$$

gdzie a_i – współczynniki modelu

e_t – wartości losowe o rozkładzie normalnym

p – rząd procesu.

Procesem ruchomej średniej rzędu q nazywamy sekwencję wartości zdefiniowanych jako:

$$(11) \quad x_t = e_t - b_1 e_{t-1} - b_2 e_{t-2} - \dots - b_q e_{t-q},$$

gdzie b_i – współczynniki modelu

e_i – wartości losowe o rozkładzie normalnym

q – rząd procesu.

Modele losowe, autoregresji i ruchomej średniej można połączyć w jednym modelu, określanym jako ARMA¹⁵ (oznaczenia jak wyżej):

$$(12) \quad x_t = a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_p x_{t-p} + e_t - b_1 e_{t-1} - b_2 e_{t-2} - \dots - b_q e_{t-q}.$$

Typ modelu można określić na podstawie kształtu funkcji autoregresji (ACF) i autoregresji cząstkowej (PACF, nie była tutaj definiowana). Funkcje te mogą łągodnie opadać albo urywać się raptownie na określonym odstępnie (por. wzory 6–9). Kształty ACF i PACF dla konkretnej próbki pozwalają wybrać najlepszy model (por. Pawłowski 2001: 71). Warto jednak podkreślić, że w badaniach tekstu i języka najczęściej do podjęcia decyzji o estymacji procesu i interpretacji danych wystarcza analiza kształtu funkcji ACF. Taki właśnie przypadek zachodzi w przedstawionej wyżej analizie *Kroniki Dalimila*.

¹⁵ Jeżeli występuje trend, stosuje się model ARIMA, jednak tutaj taki przypadek nie zachodzi.