

Marek Osiewicz, Krzysztof Skibski
Uniwersytet im. Adama Mickiewicza w Poznaniu

Struktura statystyczna polskiego systemu graficznego. Analiza historyczno-porównawcza

1. Cel i ramy teoretyczne analizy

Celem niniejszego artykułu jest statystyczno-porównawcza analiza czterech systemów pisownianych reprezentujących różne etapy rozwoju polskiego systemu graficznego. Stanowi on pierwszą próbę przeniesienia problematyki dotyczącej funkcjonalnej wartości innowacji graficznych na grunt bardziej obiektywny, niezakotwiczony w tradycyjnej, teleologicznej interpretacji rozwoju polskiej grafii¹. Korzystając z metod kwantytatywnych, analiza niniejsza wpisuje się w ten nurt badań językowych, w którym istotną rolę odgrywa sekwencyjna analiza tekstu². Jej szczegółowe hipotezy badawcze sprowadzić można do jednego, ogólnego stwierdzenia, że linearne następstwo poszczególnych jednostek językowych w tekście stanowi realizację jakiegoś procesu stochastycznego, a więc nie ma charakteru losowego³. W analizach tego typu cechą relewantną i tym samym podlegającą modelowaniu matematycznemu jest porządek tworzących go jednostek. Modelowy charakter niniejszej analizy decyduje o jej przynależności do lingwistyki systemowej⁴.

¹ W zasadzie we wszystkich dotychczasowych opracowaniach dotyczących dziejów polskiej grafii używa się wartościujących określeń typu „wzrost funkcjonalności”, „udoskonalanie”, „dziwaczna pisownia” czy „bałamutna pisownia”. Zob. np. Z. Klemensiewicz, *Historia języka polskiego*, Warszawa 1999, s. 94–99, 358–370; S. Jodłowski, *Losy ortografii polskiej*, Warszawa 1979, s. 18; S. Rospond, *Gramatyka historyczna języka polskiego*, Warszawa 1971, s. 43, 52–53; M. Osiewicz, *Czy polskiej pisowni potrzebna jest ortografia?*, „Polonistyka” 2007, nr 9, s. 18–23; T. Lisowski, *Pisownia polska. Główne fazy rozwoju (propozycja rozdziału podręcznika do nauczania treści historycznojęzykowych na studiach I stopnia)*, „Kwartalnik Językoznawczy” 2010, nr 3–4 (3–4), www.kwartajez.amu.edu.pl, s. 122 i n.

² Zob. A. Pawłowski, *Metody kwantytatywne w sekwencyjnej analizie tekstu*, Warszawa 2001, s. 23–26.

³ Ibidem, s. 25.

⁴ Zob. ibidem.

2. Podstawowe parametry analizy rozkładu liter i diad literowych

Analiza rozkładu (dyspersji) liter i diad literowych pozwala stwierdzić, w jaki sposób przekształcenia polskiego systemu graficznego wpływały na jego strukturę statystyczną. Dzięki analizie tej możliwa jest ocena stopnia zróżnicowania poszczególnych systemów pisownianych; jej podstawą jest badanie wzajemnego stosunku elementów (liter i diad literowych) częstych i rzadkich. Podstawą ogólnej segmentacji statystycznej liter i diad literowych o różnych częstościach jest współczynnik F , oznaczający średnią częstość występowania elementów (liter – F_1 , diad – F_2). Współczynnik ten oblicza się, dzieląc ogólną liczbę wystąpień liter (U_L) bądź diad literowych (U_D) w danej próbie tekstowej przez ich liczbę (L_L lub L_D). Elementy o frekwencji wyższej niż F to jednostki częste, zaś elementy o frekwencji niższej niż F to jednostki rzadkie. Im więcej w tekście rzadkich liter bądź diad literowych, tym jego struktura statystyczna jest bardziej zróżnicowana. Zastosowana analiza rozkładu liter i dwuelementowych połączeń liter wykorzystuje znacznie bardziej szczegółowe segmentacje frekwencyjne tych elementów; istotną rolę w tym procesie badawczym odgrywa też rozróżnienie między liczbą elementów (L) a ich użyciem (U).

3. Podstawowe miary teorii informacji⁵

Alternatywne, dotąd niewykorzystane w aspekcie historycznym narzędzie badań przynosi ilościowo-probabilistyczna teoria informacji⁶. Proponowana przez nią metoda obliczeń dwóch najważniejszych wielkości: entropii i redundancji pozwala w sposób obiektywny, bo niezależny od jakiegokolwiek graficznego punktu odniesienia, zmierzyć średnią pojemność informacyjną tekstu zapisanego w danym systemie pisownianym i na tej podstawie ocenić stopień zachodzących pomiędzy poszczególnymi etapami rozwojowymi zmian. Analizy tego typu nie wyczerpują oczywiście wszystkich zagadnień związanych z funkcjonalnością systemów graficznych, jednak dają dość efektywne narzę-

⁵ O pojęciu informacji i jej miarach statystycznych zob. np.: M. Jurkowski, *Teoria informacji a lingwistyka*, „Poradnik Językowy” 1965, z. 1, s. 4–16; B. Sigurt, *Struktura języka. Zagadnienia i metody językoznawstwa współczesnego*, Warszawa 1975, s. 159–167; R. Hammerl, J. Sambor, *Statystyka dla językoznawców*, Warszawa 1990, s. 361–451; K. Kunysz, *Elementy teorii informacji*, Rzeszów 1990, s. 29–35; A. Pawłowski, op.cit., s. 38–42.

⁶ O różnych odmianach (postaciach) teorii informacji zob.: M. Lubański, *O pojęciu informacji*, „Studia Philosophiae Christianae” 1974, t. 10; idem, *Filozoficzne zagadnienia teorii informacji*, Warszawa 1975; W. Nawrocki, *W poszukiwaniu istoty informacji*, w: *Analiza pojęcia informacji*, red. J.J. Jadacki, Warszawa 2003, s. 37–62.

dzie oceniania przede wszystkim istotności różnic występujących między systemami, zwłaszcza tymi, które służą do zapisywania tego samego języka.

Jądrem ilościowo-probabilistycznej teorii informacji jest założenie, że pomiędzy informacją a stopniem uporządkowania (w znaczeniu prawdopodobnościowym) systemu znakowego istnieje odwrotnie proporcjonalna zależność: im mniejsze prawdopodobieństwo sygnału, tym większą ilość informacji on niesie. Ilość informacji jest tu zatem związana z pojęciem różnorodności, nie zaś z pojęciem wiedzy. Wielkością określającą średnią ilość informacji zawartą w sygnale danego kodu jest entropia. Oblicza się ją, mnożąc prawdopodobieństwo wystąpienia każdej jednostki kodu przez jego logarytm (w naszej pracy jest to logarytm o podstawie 2, tzw. logarytm binarny lub dwójkowy), a następnie sumując poszczególne wyniki. Tak obliczana entropia zwana jest entropią bezwzględną rzędu pierwszego (H_1), gdyż podstawę jej obliczania stanowią sygnały rozpatrywane jako niezależne od siebie. Rodzajem entropii bezwzględnej jest też entropia maksymalna (H_0), charakteryzująca taki stan układu, w którym wystąpienie wszystkich jednostek cechuje jednakowe prawdopodobieństwo; jest to jednocześnie stan największej teoretycznie możliwej pojemności przekaźnikowej kodu znakowego. Oblicza się ją najprościej, gdyż wartość jej jest równa wartości logarytmu binarnego z liczby znaków współtworzących dany system ($\log_2 N$). Obie entropie bezwzględne (maksymalna i rzędu pierwszego) mają dość ograniczoną zdolność charakteryzowania nośności informacyjnej tekstów pisanych danym kodem – nie uwzględniają bowiem zależności występowania poszczególnych sygnałów (znaków) od obecności innych znaków. Stanowią jednak wielkości niezwykle ważne, gdyż są niezbędne do obliczeń innych, istotniejszych pod względem interpretacyjnym parametrów teorii informacyjnych.

W przypadku kodów językowych prawdopodobieństwo pojawienia się danej jednostki jest warunkowe, tzn. uzależnione od jednostek poprzedzających. Wartością uwzględniającą tę własność struktury przekazu jest entropia warunkowa. Wiąże ona oba typy prawdopodobieństwa jednostek: prawdopodobieństwo ogólne i prawdopodobieństwo warunkowe. W zależności od głębokości badanego związania przekazu wyróżnia się entropie rzędu drugiego (H_2), trzeciego (H_3), czwartego (H_4) itd. Ideałem obliczeniowym jest entropia rzędu najwyższego (nieskończonego), nazywana też entropią faktyczną lub graniczną danego kodu; z przyczyn natury obliczeniowej (zob. niżej) stanowi ona ideał właściwie nieosiągalny. W obliczaniu entropii warunkowych korzysta się z pomocy tzw. entropii pomocniczych (H^2 , H^3 itd.), uzyskiwanych przez mnożenie prawdopodobieństwa każdej jednostki analizowanego rzędu (diady, triady, tetrazy itd.) z jego logarytmem binarnym, a następnie przez sumowanie uzyskanych w ten sposób wyników. Na koniec wartość tę odejmuje się od

wartości entropii poprzedzającego rzędu (czyli np. entropię pomocniczą rzędu drugiego odejmuje się od entropii bezwzględnej rzędu pierwszego; wynik obliczenia jest entropią warunkową rzędu drugiego).

Istotna dla badań nad przekształceniami systemów graficznych jest też wartość wynikająca ze stosunków pomiędzy poszczególnymi entropiami, tzw. entropia względna (H_w). Wartość tę uzyskuje się, dzieląc entropię warunkową danego rzędu przez entropię maksymalną (H_0). Pojęcie entropii względnej (obliczanej osobno dla poszczególnych rzędów związania) jest najistotniejsze, podsumowujące określony etap obliczeń i – co najważniejsze – podlegające konkretnym przełożeniom interpretacyjnym. Jest ona miarą stopnia swobody wyboru danych jednostek, miarą stopnia braku zależności między poszczególnymi elementami układu. Im wyższa jej wartość, tym dany system znaków jest na poziomie realizacji tekstowej mniej związany, zaś wraz z jej spadkiem rośnie wpływ znaku poprzedniego na prawdopodobieństwo pojawienia się znaku następnego. Tak charakteryzowana entropia względna ma również swoją interpretacyjną przeciwwagę w postaci redundancji⁷. Wyrażana wzorem $1 - H_w$, jest miarą stopnia związania strukturalnego, uzależnienia swobody wyboru elementu kodu od jego poprzednika, innymi słowy: jest miarą głębi strukturalnej danego układu – rośnie ona wraz ze wzrostem wartości redundancji.

Między wartościami entropii poszczególnych rzędów istnieje stała zależność: wartość ta maleje wraz ze wzrostem obliczanego rzędu. Entropia rzędu drugiego jest zatem niższa od entropii rzędu pierwszego, entropia rzędu trzeciego – niższa od entropii rzędu drugiego itd. Z tej zależności wynika prosty wniosek, że między wartościami entropii rzędów początkowych (np. pierwszego i drugiego) a wartością entropii rzędu najwyższego (faktualną) istnieje znaczna różnica (por. np. wielkości kolejnych entropii liter języka angielskiego: $H_0 = 4,76$; $H_1 = 4,03$; $H_2 = 3,32$; $H_8 = 1,87$; $H_\infty = 1,40^8$). Niestety, uzyskiwanie entropii wyższych rzędów napotyka spore trudności. Podstawową z nich jest trudność stosowania zaprezentowanej wyżej tzw. metody mechanicznej, wymagającej wyznaczenia w określonej próbie tekstowej frekwencji wszystkich jednostek należących do obliczanego poziomu. Nie przysparza większych kłopotów obliczanie entropii rzędu pierwszego i drugiego, problemem jednak staje się uzyskanie danych frekwencyjnych potrzebnych do wyznaczenia entropii rzędu trzeciego: w przypadku kodu trzydziestoelementowego wymaga to obliczenia frekwencji aż 27 000 triad (dla entropii wyższych liczba ta rośnie: 810 000 tetrad, 24 300 000 połączeń pięcioelementowych itd., zgodnie

⁷ Jest to przeciwieństwo doskonałe, wyrażające się wzorem: $R_x + H_x = 1$ (gdzie x oznacza rząd obliczanych wielkości).

⁸ Zob. M. Jurkowski, op.cit., s. 52–53.

z zasadą, że liczba połączeń n -sygnałowych kodu liczącego s znaków wynosi s^n). Wynika stąd, że do obliczenia entropii bliskich rzędu najwyższego (w miarę wzrostu stopnia obliczeń ich wartość maleje i w końcu się stabilizuje) potrzebne jest wyznaczenie frekwencji więcej niż miliarda sygnałów. Uciążliwość metody mechanicznej wzmaga też fakt, że na poziomie tekstu wykorzystywana jest zwykle nikła część teoretycznie możliwych połączeń (np. spośród wszystkich 2 560 000 możliwych tetrad fonemowych w tekstach pojawia się zaledwie około 1,4% czwórek⁹), z czego wynika, że ogromna część pracy obliczeniowej jest daremna, a tym samym metoda mechaniczna okazuje się metodą w wysokim stopniu nieekonomiczną. Z tego powodu w obliczaniu entropii wyższych rzędów stosuje się metody eksperymentalne, polegające na odgadywaniu przez odpowiednio dobranego użytkownika danego kodu kolejnych jednostek tekstu (np. liter albo fonemów). Liczba dawanych przez niego odpowiedzi stanowi wielkość określającą warunkowe prawdopodobieństwo pojawienia się danego elementu w określonym ciągu znaków. Metoda ta, wymagająca co prawda licznych doświadczeń z udziałem większej liczby osób badanych, jest jednak znacznie mniej uciążliwa niż metoda eksperymentalna. Jej wadę stanowi mniejsza pewność uzyskiwanych w ten sposób wyników – pod tym względem metoda mechaniczna, odznaczająca się matematyczną ścisłością, jest bezkonkurencyjna¹⁰.

4. Podstawa materiałowa

Podstawę materiałową prezentowanej analizy tworzą następujące próby tekstowe: fragment *Modlitewnika Olbrachta Gasztolda*¹¹ (rękopis, część pisana pierwszą ręką: k. 5v–128v, około 1530 roku; dalej: *Gaszt*), fragment *Ksiąg o gospodarstwie domowym*¹² Piotra Krescencjusza (druk, fragment części I tekstu – szpalty 189–260 – z około 1538 roku¹³; dalej: *Cresc*), cały tekst

⁹ W. Jassem, P. Łobacz, *Fonotaktyczna analiza mówionego tekstu polskiego*, „Biuletyn Polskiego Towarzystwa Językoznawczego” 1974, t. 32, s. 186.

¹⁰ Opisy metod niemechanicznych, eksperymentalnych zob. w: R. Hammerl, J. Sambor, op.cit., s. 381–396.

¹¹ Korzystano z wydania (transliteracja i transkrypcja): A. Augustyniak, „*Modlitewnik Olbrachta Gasztolda*”. *Edycja tekstu* (praca magisterska napisana pod kierunkiem prof. W. Wydry w IFP UAM, Poznań 2006), s. 21–22 (za zgodą autorki i promotora pracy).

¹² *Piotra Creścentyna Księgi o gospodarstwie, y o opatrzeniu rozmnożenia rozlicznych pożytkow, każdemu fjanowi potrzebne*. Korzystano z egzemplarza będącego w zbiorach Biblioteki Kórnickiej (mikrofilm, sygn. Mf 4935).

¹³ Mimo że na karcie tytułowej widnieje data 1549, Kazimierz Piekarski, opierając się na analizie materiału typograficznego używanego w tłoczni Unglerów, zwrócił uwagę na fakt,

Pętli Marka Hłaski (1956; dalej: *Pętla*) oraz zbiór zapisów rozmów internetowych przeprowadzonych z użyciem komunikatora Gadu-Gadu (materiał z 2008 roku; dalej: *GG*). Zebrane próby tekstowe reprezentują różne systemy graficzne. Pisownia *Gaszt* poświadcza grafie zakorzenioną jeszcze w tradycji średniowiecza; świadczą o tym: zapisywanie samogłosek nosowych za pomocą znaków „ą” = [ɛ] i „a” = [ɛ], [a]; oznaczanie spółgłoski [v] za pomocą liter „w”, „v” i „u”, a także pisownia głoski [ś] jako „fz”¹⁴; ogólnie rzecz ujmując, jest to jeszcze w dużej mierze grafia niediakrytyczna (prócz „y”, „ÿ”, „ą”). *Cresc* prezentuje grafie typową dla druków pierwszej połowy XVI wieku, wykorzystującą już na szeroką skalę diakryty (znaki „ż”, „ć”, „ą”, „ę”, „á”, „ł”) ze skłonnością do ich nadużywania (zapisy typu *wfćżepić*), z nowoczesną pisownią nosówek „ę” // „ą”, „ł” // „l”; pisownia tego zabytku częściowo zdradza jednak jeszcze pewne przyzwyczajenia pochodzące z epoki wcześniejszej (np. „th” = [t]). *Pętla* reprezentuje standardową pisownię współczesną z drugiej połowy XX wieku (jej charakterystyka jest zbędna – jest to grafia, którą wykorzystano też do zapisania niniejszego artykułu), zaś *GG* poświadcza tzw. współczesną grafie niestandardową, innowacyjną grafie mediów elektronicznych; do jej cech charakterystycznych należą: fakultatywność stosowania znaków diakrytycznych, skłonność do przestawień połączeń literowych (tzw. literówki) oraz silna anglicyzacja.

Zgromadzony wybór tekstów nie spełnia większości wymogów stawianych przez porównawczą analizę statystyczną. Relacja proporcjonalności w zakresie stosunku objętości zgromadzonych prób tekstowych jest nierównomierna (w literach ze spacjami: *Gaszt* – 72 846 znaków, *Cresc* – 94 885 znaków, *Pętla* – 99 414 znaków, *GG* – 98 040 znaków) – spełniają ją próby: *Cresc*, *Pętla* i *GG*; *Gaszt* poświadcza ponad 20 000 (21–22%) znaków mniej niż pozostałe próby. Podkreślić jednak należy, że objętość analizowanych tekstów nie spełnia też wymogów reprezentatywności (zbyt małe próby). Pod względem tematyczno-stylistycznym teksty te nie tworzą spójnej grupy: *Gaszt* jest

że w zakresie rysunku liter tekst *Ksiąg...* przedstawia się niejednolicie: składki A–Q włącznie (szpalty 1–384) wykazują cechy znane z druków z 1538 roku, natomiast składki wstępne (strona tytułowa, epistoła wydawcy i registr) oraz składki od R (od szpalty 381/2 – na styku obu części tekstu szpalty są błędnie ponumerowane i cofnięte w numeracji o 4) aż do końca tekstu (czyli do składki Ii) prezentują stan znany z *Kroniki Bielskiego* z 1551 roku (K. Piekarski, *Miscellanea bibliograficzne*, „Przegląd Biblioteczny” 1930, t. 4, z. 4, s. 423–424). Stan ten Piekarski wytłumaczył następująco: „składki A–Q wytłoczono około 1538 roku i z jakichś bliżej niezanych nam powodów dalszego druku zaniechano; dopiero zaś w dziesięć lat później, w 1549 r., podjęto i wykończono wydanie tego dzieła. Inne hipotezy nie wydają mi się prawdopodobne”. Ibidem, s. 415–443.

¹⁴ Zob. A. Augustyniak, op.cit., s. 21–22.

tekstem religijnym, *Cresc* to podręcznik gospodarstwa wiejskiego¹⁵, *Pętla* to utwór literacki (artystyczny), zaś zapisy rozmów *GG* to w istocie pisana wersja polszczyzny konwersacyjnej (potocznej). Charakterystyka bazy materiałowej pokazuje, że wielkości statystyczne uzyskane dzięki jej analizie nie mogą być podstawą wiarygodnych wniosków: ich wartość może mieć charakter jedynie ilustracyjny, ulegną one z pewnością weryfikacji w wyniku analiz dokonanych na bardziej reprezentatywnej grupie źródeł. Tym samym charakter niniejszego artykułu określić należy jako wstępny – istotną jego część stanowią zagadnienia natury metodologicznej.

Podstawą przeprowadzonych obliczeń były frekwencje pojedynczych liter i ich diad; spację potraktowano jako znak równoważny literom (w niniejszym artykule terminy „znak” i „litera” są stosowane zamiennie, czyli jako zakresowo tożsame – odnoszą się zarówno do liter, jak i do spacji). Wielkość liter, a także wszystkie znaki interpunkcyjne uznano za cechy nieistotne. W przypadku tekstów staropolskich uwzględniono rozróżnienie między tzw. długim „r” i okrągłym „s” – ze względu na nie do końca wyjaśnioną funkcję tej opozycji oba znaki uznano za osobne jednostki. Rozróżnienia takiego nie przeprowadzono w przypadku występujących w obu XVI-wiecznych tekstach dwóch typów litery „r”: „r” zwykłego i „r” okrągłego – opozycja ta w tekstach staropolskich miała bowiem charakter – jak się zdaje – wyłącznie graficzny¹⁶. W przypadku tekstów historycznych pod uwagę wzięto wyłącznie całe słowoformy; nie uwzględniono form zniekształconych przez zewnętrzne, fizyczne uszkodzenie tekstu; pozostawiono jednak formy zawierające tzw. literówki – zdaniem piszących te słowa stanowią one immanentną część struktury stochastycznej tekstu, w którym zostały poświadczone. Ze wszystkich fragmentów usunięto cytaty z języków obcych: łacińskiego i angielskiego; w tekstach *GG* pozostawiono jednak formy typu *ok*, *sorry* jako zadomowione, mające status form zasymilowanych, pominięto natomiast obce nazwy własne.

5. Uwagi metodologiczne

Zastosowanie metod statystycznych do badań nad przekształceniami systemu graficznego sprowadza się do porównywania liczby i częstości wy-

¹⁵ W.R. Rzepka, W. Wydra, *Chrestomatia staropolska. Teksty do roku 1543*, Wrocław 1984, s. 376.

¹⁶ Tzw. „r” okrągłe występowało w zasadzie wyłącznie po literach z „brzuszkami”, np. po „o”, „b”, „d”, co jest ściśle związane z genezą tego znaku powstałego na potrzeby ligatury „or”, zob. W. Semkowicz, *Paleografia łacińska*, Kraków 2011, s. 261–262, 322. O kontekście występowania „r” okrągłego zob. też: M. Juda, *Pismo drukowane w Polsce XV–XVIII wieku*, Lublin 2001, s. 74, 90, 105, 124.

stępowania liter oraz ich połączeń, poświadczonych w poszczególnych – reprezentatywnych dla danych typów pisownianych – próbach tekstowych. Przyjęta w niniejszej propozycji badawczej perspektywa dynamiczna analizy (historyczno-porównawcza) wiąże się z dość sporą grupą problemów natury metodologicznej.

Przy analizie tekstów historycznych jedyną metodą uzyskiwania danych frekwencyjnych jest metoda mechaniczna – przesądza o tym brak możliwości odwołania się do kompetencji użytkowników dawnych systemów graficznych. Konieczność stosowania metody mechanicznej wiąże się z sygnalizowanymi już ograniczeniami natury przeliczeniowej. Mimo że do ekscerpcji i obliczeń wykorzystano komputery, to jednak praca ta w dużej mierze miała charakter ręczny (digitalizacja tekstów dawnych, wpisywanie zadanych parametrów do wyszukiwarki tekstowej, tworzenie zestawień kalkulacyjnych). Fakt ten spowodował ograniczenie pracy ekscerpcyjnej do przeliczania frekwencji poszczególnych liter i ich diad, co umożliwiło obliczenie najbardziej podstawowych wartości teoriiinformacyjnych: entropii maksymalnej oraz entropii najniższych rzędów (I i II) oraz odpowiadających im entropii względnych (w tym także wielkości im przeciwnych, czyli redundancji). Pozostaje mieć nadzieję, że to czysto techniczne utrudnienie uda się w przyszłości przekroczyć.

Istotne ograniczenia wynikają też ze specyfiki wykorzystywanej w analizie bazy źródłowej. W sporej części składa się ona z tekstów dawnych, reprezentujących odpowiednie, historyczne etapy dziejów systemu graficznego. Ze względu na przyjętą w niniejszej propozycji badawczej perspektywę porównawczą wykorzystanie tego typu tekstów wiąże się z szeregiem specyficznych problemów natury metodologicznej.

Na pierwszym miejscu wymienić należy problem objętościowej mocy i objętościowej równoważności porównywanych tekstów. W badaniach ustalających jakiegokolwiek wartości średnie (czyli statystyczne) niebagatelne znaczenie ma liczebność (tekstowa objętość) prób; ich moc reprezentatywna rośnie wraz z objętością. W perspektywie porównawczej wymóg ten nakłada się na inny, dotyczący proporcjonalności konfrontowanych prób tekstowych, dyskwalifikujący wnioski uzyskiwane na podstawie zestawiania wyników analiz opartych na różnych objętościowo bazach materiałowych. Podsumowując: podstawa materiałowa analizy powinna składać się z jak największych objętościowo i jednocześnie proporcjonalnych grup tekstów. Problem ten zyskuje na znaczeniu zwłaszcza w kontekście badań porównujących statystyczne parametry tekstów współczesnych i dawnych. W przypadku tekstów reprezentujących współczesną pisownię, zarówno standardową, jak i niestandardową (innowacyjną?), objętość próby nie stanowi problemu: dostęp do nich jest w zasadzie nieograniczony. Inaczej jest w przypadku tekstów repre-

zentujących pisownię dawną – z uwagi na stwierdzoną partykularność norm graficznych poświadczanych przez poszczególne zabytki nie tworzą one – jak teksty współczesne – spójnego pod tym względem zbioru. Wymaga to indywidualnego traktowania każdego tekstu. Takie podejście determinuje automatycznie kwestię objętości próby – jej granice wyznaczone są przez granice tekstu. Niestety, w miarę cofania się w czasie maleje zarówno zbiór tekstów, jak i ich średnia objętość. Co prawda prócz tekstów krótkich, takich jak *Kazania świętokrzyskie* (15 709 liter) czy *Kazania gnieźnieńskie* (51 137 liter), zdarzają się też teksty znacznie większe, ale jest ich niewiele (wśród tekstów opracowanych w *Bibliotece zabytków polskiego piśmiennictwa średniowiecznego*¹⁷ objętość tylko siedmiu przekracza 100 000 liter¹⁸). Warto przypomnieć, że teksty staropolskie często są dziełem więcej niż jednej ręki pisarskiej – fakt ten musi w badaniach omawianego typu znaleźć stosowne odbicie (części pisane przez różnych pisarzy czy kopistów powinny być traktowane oddzielnie). W obliczu tych faktów swoboda doboru źródeł jest znacznie ograniczona – ich objętość musi być dobierana lub redukowana do objętości wybranych źródeł staropolskich.

Ciąży też ograniczony zakres tematyczny zachowanej spuścizny tekstowej. Frekwencje liter są nieodłącznie związane z frekwencjami głosek czy fonemów, te zaś są zależne od struktur wyższego poziomu: gramatycznych i leksykalnych. W czasie zestawiania wartości kwantytatywnych charakteryzujących litery różnych tekstów zachodzi spore prawdopodobieństwo, że różnice między nimi opisują nie tylko różnice na płaszczyźnie graficznej, ale również dzielące te teksty różnice językowe. Różnice te mogą mieć charakter systemowy, wynikający ze zmian dziejowych, jakim struktura języka podlega; z sytuacją taką możemy mieć do czynienia zwłaszcza wtedy, gdy porównujemy teksty bardzo od siebie odległe chronologicznie. Istotniejszy wpływ na wyniki entropii liter mają jednak te różnice, które wynikają z probabilistycznie odmiennej struktury tekstowej źródeł odmiennych pod względem tematycznym czy stylistycznym, tj. gramatycznym i leksykalnym. Wyniki badań nad zależnościami frekwencji zarówno fonemów, jak i liter od czynników stylistyczno-tematycznych, prezentowane w pracach Bronisława Ročławskiego¹⁹ oraz

¹⁷ *Biblioteka zabytków polskiego piśmiennictwa średniowiecznego*, Instytut Języka Polskiego Polskiej Akademii Nauk w Krakowie, Kraków 2006 (edycja elektroniczna – płyta DVD).

¹⁸ Są to: *Rozmyślanie przemyskie* – 671 513 liter; *Biblia królowej Zofii* – 599 387 liter; *Kodeks Świętosławowy* – 230 906 liter; *Psalterz puławski* – 219 947 liter; *Ortyle ossolińskie* – 207 788 liter; *Psalterz floriański* – 204 697 liter; *Ortyle magdeburskie* – 195 807 liter. Dane mają charakter względny, gdyż zostały uzyskane na podstawie transkrypcji.

¹⁹ B. Ročławski, *System fonostatystyczny współczesnego języka polskiego*, Wrocław 1981.

Lothara Hoffmanna i Rajmonda Piotrowskiego²⁰, ilustrują związek zachodzący między tymi poziomami tekstu. Stąd też szczególne znaczenie w badaniach statystycznych właściwości liter ma proporcjonalny pod względem stylistycznym i tematycznym dobór źródeł. Kwestia ta znów staje się problematyczna w przypadku tekstów dawnych, które w wyborze tego typu nie pozostawiają badaczowi zbytnej swobody. Problem zyskuje na znaczeniu, gdy weźmiemy pod uwagę ograniczenia objętościowe przy tworzeniu bazy tekstów reprezentatywnych – związki statystyczne między wspomnianymi poziomami tekstu tym są wyraźniejsze, im teksty porównywane są krótsze. Dobór odpowiedniej bazy źródłowej wymaga więc sporego namysłu i musi uwzględniać oba parametry: ilościowy i jakościowy.

Osobną, problematyczną kwestią jest ustalenie relewantnych jednostek kodu. Jeśli w przypadku współczesnych tekstów standardowych zagadnienie to nie przysparza większych trudności (dyskusyjne może być np. uznanie za jednostki relewantne wielkich liter lub przerw międzywyrazowych czy niektórych znaków interpunkcyjnych), to w przypadku zarówno tekstów dawnych, jak i współczesnych tekstów niestandardowych problem ten zyskuje na znaczeniu. Sporna jest relewancja różnych wariantów tej samej litery (np. „f” i „s”), a także dystynktywność połączeń przypadkowych (powstałych w wyniku błędu), cytatów obcojęzycznych czy skrótów. Problem stanowi już sama kwalifikacja danego zjawiska jako np. potknięcia pióra (literówki) czy nawet cytatu obcojęzycznego (*okey, sorry* itp.). Nie ulega wątpliwości, że praca ekscerpcyjna (przeliczeniowa) musi być poprzedzona odpowiednimi decyzjami prowadzącymi do „oczyszczenia” tekstu z cech nierelewantnych.

6. Parametry charakteryzujące pojedyncze litery

6.1. Liczba liter (L_1)

Analizowane fragmenty tekstów różnią się liczbą użytych znaków:

Gaszt 31 znaków: #, a, á, e, e, i, i, o, u, y, ý, ý, a, b, c, d, f, g, h, k, l, m, n, p, r, s, f, t, v, w, z;

Cresc 34 znaki: #, a, á, a, e, e, i, o, y, b, c, c, d, f, g, h, j, k, l, l, m, n, n, p, r, s, f, t, u, v, w, x, z, z;

²⁰ L. Hoffmann, R.G. Piotrowski, *Beiträge zur Sprachstatistik*, Leipzig 1979, s. 92–93, cyt. za: R. Hammerl, J. Sambor, op.cit., s. 399.

- Pętla* 33 znaki: #, a, ą, e, ę, i, o, ó, u, y, b, c, ć, d, f, g, h, j, k, l, ł, m, n, ń, p, r, s, ś, t, w, z, ź, ż;
- GG* 36 znaków: #, a, ą, e, ę, i, o, ó, u, y, b, c, ć, d, f, g, h, j, k, l, ł, m, n, ń, p, q, r, s, ś, t, v, w, x, z, ź, ż.

Największą liczbę liter wykorzystują teksty powstałe w komunikatorze Gadu-Gadu, najmniejsza liczba znaków poświadczona została w *Modlitewniku Olbrachta Gasztolda*. Pod tym względem najbliższe sobie są *Cresc* i *Pętla* (34 i 33 znaki), które dzieli od pozostałych tekstów zbliżona liczba liter (2–3).

6.2. Frekwencje liter (U_1)

Listy frekwencyjne liter poświadczonych w poszczególnych tekstach prezentują się następująco:

- Gaszt* # 10726, i 7389, a 5384, e 5356, z 5045, o 4862, f 3787, n 2829, c 2643, l 2542, r 2387, u 2100, m 2052, t 1960, k 1745, d 1684, y 1678, p 1598, w 1542, b 1205, g 912, v 820, s 781, h 585, ą 485, y 480, ý 217, f 49, á 1, ę 1, ĩ 1;
- Cresc* # 15519, i 7829, e 6960, a 6834, o 6411, z 4670, n 3800, y 3684, f 3486, r 3356, w 3288, t 3154, c 2780, d 2545, k 2536, m 2405, p 2046, ź 1983, b 1828, u 1495, l 1402, g 1324, ł 1117, ą 1061, h 869, ę 858, ć 660, á 445, j 155, v 152, s 100, ń 55, f 44, x 32;
- Pętla* # 16066, i 7484, a 7248, e 6460, o 5682, z 5285, n 3828, s 3394, y 3187, c 3163, w 3159, t 3099, r 3097, ł 2959, k 2921, d 2866, m 2842, p 2660, u 1848, j 1764, l 1546, ę 1511, b 1476, g 1118, ą 1030, ź 802, ś 778, h 772, ó 596, ć 424, f 162, ń 126, ź 61;
- GG* # 17450, a 8430, e 7987, i 6661, o 6514, z 4924, n 4151, s 3682, t 3647, c 3165, m 3005, r 2845, k 2758, w 2733, y 2619, d 2513, j 2329, p 2144, l 2040, u 1638, b 1593, ł 958, h 899, g 891, ę 573, ź 431, ś 380, ó 277, ą 268, ć 252, f 176, ń 48, ź 36, x 12, v 10, q 1.

6.3. Dyspersja liter

Wartość średniej arytmetycznej występowania pojedynczych liter (F_1) w porównywanych próbach tekstowych różni się – *Gaszt*: $F_1 = 2350$, *Cresc*: $F_1 = 2791$, *Pętla*: $F_1 = 3013$, *GG*: $F_1 = 2723$. Litery o frekwencji wyższej niż F_1 to litery częste, litery o frekwencji niższej niż F_1 to litery rzadkie. W celu porównania dyspersji liter w analizowanych tekstach w niniejszej analizie przyjęto bardziej szczegółową ich segmentację frekwencyjną, będącą zmodyfikowaną wersją segmentacji stosowanej przez Władysława Kuraskiewicza

do leksykalnej analizy statystycznej²¹. Podział ten polega na wyodrębnieniu wśród liter następujących grup frekwencyjnych:

- litery częste – 10% wszystkich znaków użytych w tekście,
- litery niezbyt częste – od poniżej 10% do średniej arytmetycznej F_1 ,
- litery rzadkie, o częstości poniżej średniej arytmetycznej, ale większej od 1,
- litery bardzo rzadkie (wyjątkowe), o częstości równej 1.

Dyspersję liter *Gaszt*, *Cresc*, *Pętli* i *GG* pod względem ich frekwencji tekstowej w wyszczególnionych grupach ilustruje tabela 1.

Tabela 1. Dyspersja znaków

	Gr. fr.	Litery				Użycia		Stosunek grup 1., 2., 3., do 4., 5., 6.			
		ranga	frekwencja	liczba (L_L)	$\%L_L$	liczba (U_L)	$\%U_L$	litery		użycia	
								liczba (L_L)	$\%L_L$	liczba (U_L)	$\%U_L$
<i>Gaszt</i>	1.	1–3	10 726–5 384	3	10,0	23 499	32,3	11	35	52 950	73
	2.	4–11	5 356–2 387	8	25,8	29 451	40,4				
	3.	12–20	2 100–1 205	9	29,0	15 564	21,4	20	65	19 896	27
	4.	21–29	912–1	11	35,5	4 332	5,9				
<i>Cresc</i>	1.	1–3	15 519–6 960	3	10,0	30 308	31,9	12	35	68 991	73
	2.	4–12	6 834–3 154	9	26,5	38 683	40,8				
	3.	13–24	2 780–1 061	12	35,3	22 522	23,7	22	65	25 892	27
	4.	25–34	869–32	10	29,4	3 370	3,6				
<i>Pętla</i>	1.	1–3	16 066–7 248	3	10,0	30 798	31,0	13	39	71 152	72
	2.	4–13	6 460–3 097	10	30,3	40 354	40,6				
	3.	14–25	2 959–1 030	12	36,4	24 541	24,7	20	61	28 262	28
	4.	26–33	802–61	8	24,2	3 721	3,7				
<i>GG</i>	1.	1–4	17 450–6 661	4	10,0	40 528	41,3	14	39	77 952	80
	2.	5–14	6 514–2 733	10	27,8	37 424	38,2				
	3.	15–21	2 619–1 593	7	19,4	14 876	15,2	22	61	20 088	20
	4.	22–36	958–1	15	41,7	5 212	5,3				

Z danych zawartych w tabeli 1. wynika, że największe różnice procentowego udziału znaków dotyczą liter rzadkich i bardzo rzadkich (grupy 3. i 4.). Zwraca uwagę zwłaszcza chronologiczny przyrost repertuaru liter rzadkich w *Gaszt*, *Cresc* i w *Pętli* (29,0% // 35,3% // 36,4%), choć zauważyć trzeba, że

²¹ W. Kuraszkiewicz, *Częstotliwość wyrazów w „Panu Tadeuszu” Adama Mickiewicza i w „Wizerunku” Mikołaja Reja*, w: *Polski język literacki. Studia nad historią i strukturą*, Warszawa 1986, s. 696–708. Nie jest to rozwiązanie nowatorskie – inny model analizy statystyczno-leksykalnej do badania rozkładu najczęstszych diad fonemowych zastosował już B. Roślowski (op.cit., s. 147–149).

w tej grupie frekwencyjnej *Cresc* i *Pętla* poświadczają zbliżony stan, odbiegający znacznie od stanu charakteryzującego *Gaszt*. W przypadku tych trzech prób tekstowych tendencja chronologiczna, lecz o odwróconym kierunku, charakteryzuje też procentowy udział znaków z 4. grupy frekwencyjnej (*Gaszt* 35,5% // *Cresc* 29,4% // *Pętla* 24,4%). W obrębie liter niezbyt częstych zwraca uwagę stosunkowo liczny w porównaniu z pozostałymi próbami tekstowymi repertuar znaków poświadczonych w *Pętli* (30,3% // *Gaszt* 25,8% // *Cresc* 26,5%). Wskazania uzyskane dla liczby pojedynczych znaków pochodzących z tych trzech prób tekstowych świadczą o różnorodności liter w tekście reprezentującym grafie późnośredniowieczną (najwyższy odsetek liter o frekwencji równej 1); najwyższą przewidywalnością na tym poziomie odznacza się współczesna grafia standardowa (najniższy odsetek znaków o frekwencji równej 1 i najwyższy odsetek znaków należących do 2. grupy frekwencyjnej).

Na tym tle zupełnie różną dyspersją na poziomie liczby liter charakteryzuje się próba *GG*. W stosunku do pozostałych prób największy kontrast przyjmuje ona w zakresie znaków należących do 3. i 4. grupy frekwencyjnej. W obrębie liter rzadkich w *GG* odnotowano najniższy odsetek jednostek (19,4%), w grupie 4. zaś próba ta poświadcza najwyższą wartość (41,7%). Świadczy to o znacznej różnorodności (nieprzewidywalności) znaków wykorzystanych w tekstach wytworzonych z użyciem komunikatora Gadu-Gadu.

W zakresie użyc liter różnice między próbami *Gaszt*, *Cresc* i *Pętla* nie są już tak wyraźne. Odnotować należy jedynie nieznaczną chronologiczną tendencję wzrostową odsetka użyc liter rzadkich (*Gaszt* 21,4% // *Cresc* 23,7% // *Pętla* 24,7%); pozostaje ona w zgodzie z odnotowanym w tych trzech próbach wzrostem odsetka od liczby znaków należących do 3. grupy frekwencyjnej. Zwraca też uwagę zgodny w *Cresc* i w *Pętli* (*Cresc* 3,6%, *Pętla* 3,7%) spadek użyc liter z grupy ostatniej wobec wartości wskaźnika odnotowanego dla próby *Gaszt* (5,9%).

Dyspersja użyc pojedynczych liter w *GG* różni się od pozostałych prób większością wskazań. W próbie tej jedynie znaki należące do 2. grupy frekwencyjnej charakteryzują się zbliżonym do pozostałych prób tekstowych pułapem użyc (choć nieznacznie niższym: 38,2% // *Gaszt* 40,4%; *Cresc* 40,8%; *Pętla* 40,6%). W grupie 1. w *GG* odnotowano znacznie wyższy procent użyc liter niż w pozostałych próbach (41,3% // *Gaszt* 32,3%; *Cresc* 31,9%; *Pętla* 31,0%). W zakresie użyc znaków z grupy 4. próbę tę cechuje wskaźnik zbliżony do wskaźnika obliczonego dla *Gaszt*; owa relacja pozostaje w zgodzie ze zbliżonym w obu tekstach zestawem liter należących do tej grupy frekwencyjnej. W obrębie grupy 3. natomiast w tej próbie tekstowej litery wystąpiły o wiele rzadziej niż w próbach pozostałych (*GG* 15,2% // *Gaszt* 21,4%; *Cresc* 23,7%; *Pętla* 24,7%). W zakresie dyspersji użyc pojedynczych liter *GG* cha-

rakteryzuje się zatem przede wszystkim wysokimi wskaźnikami obliczonymi dla znaków ze skrajnych grup frekwencyjnych. Świadczy to o przeciwstawnych tendencjach współczesnej grafii niestandardowej: z jednej strony odznacza się ona najwyższymi ze wszystkich przebadanych prób wskaźnikami obliczonymi dla liter częstych (jako jedyna próba zaświadcza w tej grupie 4., a nie 3 znaki), z drugiej zaś – charakteryzuje się wysokimi wartościami uzyskanymi dla liter bardzo rzadkich.

W zakresie wskazań obliczonych dla zbiorów liter częstych i rzadkich (czyli znaków z grup 1. i 2. oraz 3. i 4.) teksty współczesne przeciwstawiają się tekstom dawnym – w *Pętli* i w *GG* poświadczony został identyczny, wyższy niż w próbach pochodzących z tekstów XVI-wiecznych odsetek liter częstych (*Pętla* i *GG*: 39% znaków częstych, 61% znaków rzadkich // *Gaszt* i *Cresc*: 35% znaków częstych, 65% znaków rzadkich). W zakresie odsetka użyć liter częstych i rzadkich wyróżniają się jedynie wartości uzyskane dla *GG* (80% użyć znaków częstych, 20% użyć znaków rzadkich) – w pozostałych tekstach wartości te są niższe i nieomal identyczne (*Gaszt* i *Cresc*: 73% liter częstych, 27% liter rzadkich; *Pętla*: 72% liter częstych, 28% liter rzadkich). Obserwacje te świadczą o zaznaczającej się we współczesnych systemach graficznych tendencji do zmniejszania różnorodności na poziomie repertuaru znaków oraz o zaznaczającej się w rozwiązaniach właściwych współczesnej grafii niestandardowej tendencji do stosowania liter o najniższych frekwencjach (niższy niż w *Pętli* udział użyć liter rzadkich w *GG* przy tak samo procentowo licznych ich zbiorze świadczy o większym udziale w *GG* znaków o frekwencji równej 1; zob. też wskazania dla liter z grupy 4. obliczone dla *GG*).

6.4. Entropie liter

Podstawowe parametry teorii informacyjnej obliczone dla pojedynczych liter poświadczonych w analizowanych próbach tekstów przyjmują następujące wartości (tabela 2.):

Tabela 2. Entropie liter (pierwszego rzędu): bezwzględna, względna; redundancja względna liter

	<i>Gaszt</i>	<i>Cresc</i>	<i>Pętla</i>	<i>GG</i>
H_0	4,95	5,09	5,04	5,17
H_1	4,29	4,39	4,47	4,31
H_w^1	0,87	0,86	0,89	0,83
R_1	0,13 (13%)	0,14 (14%)	0,11 (11%)	0,17 (17%)

Wartość entropii maksymalnej (H_0) wynosi: dla *Gaszt* – 4,95 bita na znak; dla *Cresc* – 5,09 bita na znak; dla *Pętli* – 5,04 bita na znak; dla *GG* – 5,17 bita na znak. Wartości te oznaczają maksymalną zdolność informacyjną kodów graficznych w sytuacji pełnej niezależności występowania wszystkich znaków (gdy każdy znak jest jednakowo prawdopodobny). Uwzględniająca niezależne prawdopodobieństwo znaków entropia rzędu pierwszego przyjmuje wartości niższe (H_1 : *Gaszt* – 4,29; *Cresc* – 4,39; *Pętla* – 4,47; *GG* – 4,31) i zmienia wzajemne zależności między poszczególnymi próbami tekstowymi: najwyższą wartość informacyjną ma w tym zestawieniu *Pętla*, choć w zakresie wartości entropii maksymalnej sytuowała się dopiero na trzecim miejscu (po *GG* i *Cresc*). Stosunki między wartościami zmieniają się po uwzględnieniu wzajemnych relacji między obiema entropiami. Najwyższą entropią względną rzędu pierwszego (H_w^1) odznacza się *Pętla* (0,89), średnie wartości ten parametr przyjmuje w *Gaszt* (0,87) i *Cresc* (0,86), zaś najniższe – w *GG* (0,83).

Z danych zawartych w tabeli 2. wynika, że w zakresie wartości entropii względnej analizowane teksty tworzą trzy odrębne grupy. Jedną z nich, o najwyższej wartości entropii względnej, stanowi *Pętla*; tekstom XVI-wiecznym właściwe są znacznie niższe i jednocześnie zbliżone wielkości tego parametru. Najniższą wartość entropia względna rzędu pierwszego przyjmuje dla tekstu *GG*. Z danych tych wynika, że na tym poziomie obliczeniowym kodem najskuteczniej wyzyskującym swe możliwości informacyjne jest współczesny, skodyfikowany system graficzny reprezentowany przez *Pętlę*. W zakresie niezależnego prawdopodobieństwa występowania liter charakteryzuje się on tym samym największą swobodą wyboru i najmniejszym stopniem strukturalnego związania (redundancja: 11,4%). W pozostałych tekstach możliwości informacyjne kodu graficznego wykorzystywane są znacznie słabiej; rozkład frekwencji liter w tych tekstach cechuje się tym samym, ale większym niż w przypadku *Pętli* stopniem związania strukturalnego (*Gaszt* – 13,5%; *Cresc* – 13,7%; *GG* – 16,6%). Zaskakuje jednak to, że *Gaszt* cechuje się nieco wyższym wskaźnikiem entropii względnej niż *Cresc*. Oba teksty są reprezentantami tradycji uchodzących za opozycyjne pod względem funkcjonalności – *Cresc* jako druk wczesnorenesansowy poświadcza wstępny etap dynamicznego procesu normalizowania się polskiej pisowni; wyniki powyższego zestawienia pokazują, że na poziomie frekwencji liter tekst ten charakteryzuje się podobnym, a nawet odrobinę niższym zakresem wykorzystania tkwiących w systemie możliwości informacyjnych niż zanurzony jeszcze w tradycji piśmiennej średniowiecza tekst *Modlitewnika Olbrachta Gasztołda*.

ep, er, es, ef, et, eu, ew, ey, ez, eż, e#, eb, ec, ec, ed, eg, ek, ep, ef, et, ew, ez, eż, f#, fe, fi, fl, fo, fr, g#, ga, ga, gd, ge, gi, gl, gl, gm, gn, go, gr, gu, gw, gy, h#, ha, ha, hb, hc, he, he, hi, hk, hl, hl, hm, hn, ho, hr, hf, ht, hu, hw, hy, hż, i#, ia, ia, ia, ib, ic, ic, id, ie, ie, if, ig, ii, ij, ik, il, il, im, in, io, ip, ir, is, if, it, iu, iw, ix, iy, iz, iz, j#, ja, jd, je, jh, jm, jn, jo, jr, js, jf, k#, ka, ka, kc, ke, ke, ki, kk, kl, kl, km, kn, ko, kr, kf, kt, ku, kw, kx, ky, kz, kz, l#, la, la, lb, lc, le, le, lg, li, lk, ll, ln, lo, lf, lu, lw, lz, l#, la, la, lc, ld, le, le, li, lk, ll, lm, ln, lo, lf, lt, lu, lw, ly, m#, ma, ma, mb, mc, md, me, me, mg, mi, mk, ml, ml, mn, mo, mp, mr, mf, mu, my, mż, n#, na, na, nc, nd, ne, ne, nf, ng, ni, nk, nn, no, nf, nt, nu, ny, nż, n#, nb, nf, o#, ob, oc, oc, od, oe, of, og, oi, oj, ok, ol, ol, om, on, on, op, or, os, of, ot, ou, ow, ox, oy, oz, oz, p#, pa, pa, pc, pe, pe, ph, pi, pj, pk, pl, pl, pn, po, pr, pf, pt, pu, py, pz, r#, ra, ra, rb, rc, rd, re, re, rg, ri, rk, rl, rl, rm, rn, ro, rp, rr, rf, rt, ru, rw, ry, rz, rz, s#, sa, sc, se, si, sk, so, sp, st, su, sz, f#, fa, fa, fc, fc, fd, fe, fi, fk, fl, fl, fm, fn, fo, fp, fr, fl, ft, fu, fw, fy, fz, fz, t#, ta, ta, tc, te, te, th, ti, tk, tl, tl, tn, to, tr, tf, tu, tw, ty, tz, u#, ua, ub, uc, ud, ue, ug, ui, uk, ul, ul, um, un, uo, up, ur, us, uf, ut, uw, uy, uz, uz, v#, va, vb, vc, vd, ve, vi, vk, vl, vl, vm, vo, vp, vr, vf, vt, vw, vy, vz, vz, w#, wa, wa, wc, wd, we, we, wi, wk, wl, wl, wn, wo, wp, wr, wf, wt, wu, wy, wz, wż, x#, xi, xu, y#, ya, yb, yc, yc, yd, ye, yg, yi, yj, yl, yl, ym, yn, yp, yr, yf, yt, yu, yw, yz, yż, z#, za, za, zb, zc, zd, ze, ze, zg, zi, zk, zl, zl, zm, zn, zo, zp, zr, zl, zt, zu, zv, zw, zy, z#, za, za, zb, zc, zd, ze, ze, zf, zg, zi, zk, zl, zl, zm, zn, zo, zp, zr, zl, zt, zu, zw, zy;

Pełta 648 diad: #a, #b, #c, #d, #e, #f, #g, #h, #i, #j, #k, #l, #l, #m, #n, #o, #ó, #p, #r, #s, #ś, #t, #u, #w, #y, #z, #ż, #ż, a#, aa, ab, ac, ac, ad, af, ag, ah, ai, aj, ak, al, al, am, an, an, ap, ar, as, as, at, au, aw, ay, az, az, az, a#, ab, ac, ac, ad, ag, ak, al, ap, as, as, at, aw, az, az, b#, ba, ba, bc, bd, be, be, bi, bj, bk, bl, bl, bn, bo, bó, br, bs, bt, bu, by, c#, ca, ca, ce, ce, ch, ci, cj, ck, cn, co, có, cp, cu, cw, cy, cz, c#, cm, d#, da, da, db, dc, dd, de, de, di, dj, dk, dl, dl, dm, dn, do, dó, dp, dr, ds, dt, du, dw, dy, dz, dż, dż, e#, eb, ec, ec, ed, ee, ef, eg, ei, ej, ek, el, el, em, en, en, eo, ep, er, es, es, et, eu, ew, ez, eż, eż, e#, eb, ec, ec, ed, eg, ek, el, el, ep, es, es, et, eż, fa, fe, fi, fl, fn, fo, fr, ft, fu, g#, ga, ga, gd, ge, ge, gi, gl, gl, gm, gn, go, gó, gr, gu, gw, gz, h#, ha, hc, he, he, hi, hl, hl, hm, hn, ho, hó, hr, hś, hu, hw, hy, i#, ia, ia, ib, ic, ic, id, ie, ie, ig, ii, ij, ik, il, il, im, in, in, io, ió, ip, ir, is, is, it, iu, iw, iz, iz, iz, j#, ja, ja, jb, jc, jd, je, je, jg, ji, jk, jl, jm, jn, jo, jp, jr, js, js, ju, jw, jż, k#, ka, ka, kb, kc, ke, ke, ki, kk, kl, kl, kn, ko, kó, kr, ks, kt, ku, kw, kz, l#, la, la, lb, lc, ld, le, le, lg, li, lk, ll, lm, ln, lo, ló, ls, lt, lu, lw, l#, la, la, lb, lc, le, le, lg, lk, ln, lo, ló, lp, ls, lt, lu, ly, lz, m#, ma, ma, mc, md, me, me, mf, mg, mi, mk, ml, ml, mn, mo, mó, mp, mr, ms, mś, mt, mu, mw, my, n#, na, na, nc, nd, ne, ne, nf, ng, ni, nj, nk, nn, no, nó, ns, nt, nu, ny, nż, n#, nc, nk, nm, ns, o#, ob, oc, oc, od, of, og, oh, oi, oj, ok, ol, ol, om, on, on, oo, op, or, os, os, ot, ou, ow, oz, oz, ób, óc, ód, óg, ój, ók, ól, ól, ór, ós, ót, ów, óz, óż, óż, p#, pa, pa, pc, pe, pe, pi, pk, pl, pl, pn, po, pó, pr, ps, pt, pu, py, r#, ra, ra, rb, rc, rc, rd, re, re, rg,

ri, rk, rl, rł, rm, rn, ro, ró, rp, rs, rś, rt, ru, rw, ry, rz, rż, s#, sa, są, sc, se, se, sf, si, sk, sł, sm, sn, so, só, sp, sr, ss, st, su, sw, sy, sz, ś#, śc, ść, śk, śl, śm, śn, śń, śp, śr, św, t#, ta, tą, tc, te, tę, ti, tk, tl, tł, tn, to, tó, tp, tr, tu, tw, ty, tz, u#, ua, ub, uc, uc, uc, uf, ug, ui, uj, uk, ul, uł, um, un, uń, up, ur, us, uś, ut, uw, uz, uż, w#, wa, wą, wb, wc, wd, we, wę, wg, wi, wj, wk, wl, wł, wm, wn, wo, wó, wp, wr, ws, wś, wt, wu, wy, wz, y#, yb, yc, yć, yd, yf, yg, yi, yj, yk, yl, ył, ym, yn, yń, yo, yp, yr, ys, yś, yt, yu, yw, yy, yz, yź, yż, z#, za, za, zb, zc, zd, ze, zę, zg, zi, zj, zk, zl, zł, zm, zn, zo, zó, zp, zr, zs, zt, zu, zw, zy, zz, ż#, źc, źd, źl, źm, źn, źr, źw, ź#, ža, ża, żb, żc, żd, że, że, żk, źl, źn, żo, żó, źr, źs, żu, ży;

GG 746 diad: #a, #b, #c, #ć, #d, #e, #f, #g, #h, #i, #j, #k, #l, #ł, #m, #n, #o, #ó, #p, #q, #r, #s, #ś, #t, #u, #v, #w, #y, #z, #ż, #ź, a#, ab, ac, ac, ad, af, ag, ah, ai, aj, ak, al, ał, am, an, ań, ao, ap, ar, as, aś, at, au, av, aw, ax, ay, az, aź, aż, a#, ab, ac, ac, ad, ag, ał, an, ap, as, aś, at, az, aż, b#, ba, ba, bc, be, be, bh, bi, bk, bl, bł, bm, bn, bo, bó, br, bs, bt, bu, by, bz, c#, ca, ca, cc, cd, ce, ce, ch, ci, cj, ck, cl, cm, cn, co, có, cp, cs, cś, ct, cu, cw, cy, cz, cz, ć#, ćc, ćm, ćn, ćw, d#, da, da, db, dc, dd, de, de, df, dg, di, dj, dk, dl, dł, dm, dn, do, dó, dp, dr, ds, dś, dt, du, dw, dy, dz, dź, dż, e#, ea, eb, ec, ec, ed, ee, ef, eg, eh, ei, ej, ek, el, eł, em, en, eń, eo, ep, er, es, eś, et, eu, ev, ew, ex, ey, ez, eź, eż, e#, eb, ec, ed, eg, ei, ej, ek, el, eł, em, en, eo, ep, es, eś, et, ez, eź, f#, fa, fe, ff, fj, fk, fl, fm, fn, fo, fr, fu, fy, fz, g#, ga, ga, gd, ge, ge, gg, gh, gi, gl, gl, gm, gn, go, gó, gr, gu, gw, gz, h#, ha, ha, hc, hd, he, he, hh, hi, hj, hl, hl, hm, hn, ho, hó, hr, ht, hu, hw, hx, hy, i#, ia, ia, ib, ic, ic, id, ie, ie, if, ig, ih, ii, ij, ik, il, il, im, in, in, io, ió, ip, ir, is, is, it, iu, iw, ix, iz, iz, j#, ja, ja, jb, jc, jd, je, je, jg, jh, ji, jj, jk, jl, jm, jn, jo, jp, jr, js, js, jt, ju, jw, jz, k#, ka, ka, kb, kc, ke, ke, ki, kj, kk, kl, kł, km, kn, ko, kó, kp, kr, ks, kś, kt, ku, kw, ky, kz, kż, l#, la, la, lb, lc, ld, le, le, li, lk, ll, lm, ln, lo, ló, lp, ls, lt, lu, lw, ly, lz, lż, l#, la, la, lb, lc, ld, le, le, li, lk, ln, lo, ló, ls, lu, ty, m#, ma, ma, mb, mc, me, me, mf, mg, mh, mi, mj, mk, ml, mł, mm, mn, mo, mó, mp, mr, ms, mś, mt, mu, my, n#, na, na, nb, nc, nd, ne, ne, nf, ng, nh, ni, nj, nk, nn, no, nó, np, ns, nt, nu, nw, nx, ny, nz, n#, nć, nć, nć, nć, nś, o#, oa, ob, oc, oc, od, oe, of, og, oh, oi, oj, ok, ol, oł, om, on, oń, oo, op, or, os, oś, ot, ou, ov, ow, ox, oy, oz, oż, ó#, óa, ób, óc, ód, óg, ój, ók, ól, ół, ór, óś, ót, ów, óz, óż, p#, pa, pa, pc, pe, pe, pi, pk, pl, pł, pm, pn, po, pó, pp, pr, ps, pt, pu, pw, py, pz, qr, r#, ra, ra, rb, rc, rd, re, re, rf, rg, ri, rk, rl, rł, rm, rn, ro, ró, rp, rr, rs, rś, rt, ru, rw, ry, rz, s#, sa, są, sb, sc, sc, sd, se, se, sf, si, sj, sk, sl, sł, sm, sn, so, só, sp, sr, ss, st, su, sw, sy, sz, ś#, śb, śc, śc, śg, śl, śł, śm, śn, śp, śr, śs, św, t#, ta, tą, tc, te, tę, tf, tg, th, ti, tj, tk, tl, tł, tm, tn, to, tó, tp, tr, ts, tt, tu, tv, tw, ty, tz, u#, ua, ub, uc, ud, uf, ug, uh, ui, uj, uk, ul, uł, um, un, up, ur, us, uś, ut, uu, uw, ux, uz, uż, v#, va, ve, vi, vo, w#, wa, wą, wc, wd, we, wę, wf, wg, wi, wj, wk, wl, wł, wm, wn, wo, wó, wp, wr, ws, wt, wu, ww, wy, wz, x#, xa, xp, xt, xu, xy, y#, ya, yb, yc, yć, yd, ye, yf, yg, yi, yj, yk, yl, ył, ym, yn, yo, yp, yr, ys, yś, yt, yu, yw, yz, yź, yż, z#, za, za, zb, zc, zd, ze, zę, zg, zi, zj,

zk, zl, zł, zm, zn, zo, zó, zp, zr, zs, zt, zu, zw, zy, zz, ż#, źc, źć, źd, źl, źm, źn, źt, ź#, źa, źą, źb, źc, źd, źe, źk, źl, źn, źo, źr, źs, źu, źy, źz.

Podobnie jednak jak w przypadku liczby liter, pod względem liczby diad literowych najbliższe sobie są *Cresc* i *Pętla* (różnica 13 diad); w tym zakresie różnice między nimi a pozostałymi dwoma tekstami są znaczniejsze (*Gaszt–Cresc*: 68 diad; *Gaszt–Pętla*: 81 diad; *GG–Cresc*: 111 diad; *GG–Pętla*: 98 diad; *Gaszt–GG*: 179 diad). Poświadczony w próbach tekstowych diady stanowią nieco ponad połowę wszystkich teoretycznie możliwych dwuelementowych połączeń znaków²²; procentowy wskaźnik ich wykorzystania we wszystkich tekstach jest zbliżony: *Gaszt* – 59%, *Cresc* – 55%, *Pętla* – 60%, *GG* – 58%. Podkreślić jedynie należy, że największy odsetek możliwych połączeń odnotowano we współczesnym tekście standardowym, zaś w stopniu najmniejszym możliwości te wykorzystane zostały w tekście drukowanym z pierwszej połowy XVI wieku.

7.2. Frekwencje diad literowych (U_D)

Listy frekwencyjne diad znakowych poświadczonych w poszczególnych tekstach prezentują się następująco:

Gaszt ie 3042, cz 2036, e# 1675, a# 1670, i# 1594, zi 1336, #i 1181, #p 1141, ni 1072, #f 947, #n 849, ze 840, na 820, ia 812, #m 735, u# 729, o# 728, rz 726, mi 683, #k 681, eł 628, z# 628, s# 626, to 625, #z 619, #b 608, #v 587, ý# 583, or 575, ch 566, łf 566, pr 565, al 543, #o 528, an 525, za 519, of 512, łz 504, #w 478, ui 475, m# 471, ac 469, zý 467, #t 457, dz 453, kt 450, go 427, le 427, ic 423, em 420, ro 409, łt 408, po 405, la 401, il 400, l# 397, oi 397, #c 393, od 387, ra 383, #d 382, mo 362, #a 356, fu 344, bo 340, łc 337, li 334, lo 334, eg 330, pa 321, lu 316, en 303, ko 301, tw 299, vi 296, rý 285, ki 278, ua 272, on 271, el 262, ad 259, al 254, h# 254, wł 252, es 249, #r 246, łl 246, wo 242, łk 237, d# 232, ec 232, ka 226, ol 226, ta 226, ou 224, n# 223, og 222, ýs 216, da 212, ed 211, ei 211, #g 209, wi 208, ai 206, ob 204, ma 202, oz 201, au 200, oc 200, am 198, a# 198, bi 191, te 189, fw 187, fe 177, zn 177, ar 175, at 175, io 175, kr 175, do 172, ow 171, ba 168, ią 168, ez 165, no 161, w# 159, fy 155, y# 154, we 153, #l 150, ek 149, eb 148, pi 148, zo 148, ne 145, zy 144, er 140, vý 140, ýc 139, dl 138, om 135, ak 134, bý 133, az 131, ný 130, #s 129, ým 128, fa 126, zw 126, aw 122, im 121, ul 121, iu 120, ku 120, hu 119, il 119, fp 118, tr 118,

²² Oblicza się ją, podnosząc do kwadratu liczbę wszystkich pojedynczych znaków użytych w danym tekście. Dla poszczególnych prób tekstowych wynoszą one: *Gaszt* – 961 diad, *Cresc* – 1156 diad, *Pętla* – 1089 diad, *GG* – 1296 diad.

gi 116, ot 116, ri 116, tk 115, ok 113, ýf 108, uc 105, zu 105, ha 103, du 102, t# 102, ýl 97, ý# 96, tÿ 95, wa 92, mu 89, ud 89, ąd 88, yl 88, zl 86, ga 84, ąt 82, dn 82, sm 81, ap 79, in 79, ýn 79, lý 78, zm 78, ov 77, ab 76, ýt 76, by 75, bą 73, zý 72, gl 71, k# 70, mÿ 70, wn 70, br 68, f# 67, as 66, bl 66, ve 66, dr 65, iw 65, rn 64, zd 64, fo 63, uo 63, ep 62, nu 62, fy 62, mn 61, eu 59, gr 55, lk 54, nf 53, pu 53, ag 52, yc 52, zb 52, ik 51, fm 51, wz 51, zk 51, nn 50, re 50, ru 50, wÿ 50, g# 49, op 49, ýi 49, me 48, ry 47, ev 46, it 46, vo 45, gm 44, vc 44, wc 44, iz 43, rc 43, ýu 43, ir 42, ýz 42, ho 41, ln 41, dÿ 40, id 40, kl 40, fn 40, ue 40, ut 40, vf 40, av 38, kd 38, rý 38, ty 38, wd 37, pl 36, rd 36, ýb 36, wr 35, de 34, ys 34, dą 33, ew 33, zr 33, rf 32, zg 32, cr 31, os 31, ug 31, ny 30, va 30, tc 29, pe 28, fi 28, wp 28, ly 27, tn 27, up 27, vk 27, zą 27, lf 26, tu 26, ąc 25, bý 25, iv 25, lm 25, ýd 25, zc 25, dk 24, gu 24, mą 24, fr 24, be 23, pf 23, rą 22, uz 22, wl 22, lc 21, nc 21, fi 21, vy 21, ýs 21, gd 20, rp 20, uk 20, ýf 20, yt 20, et 19, fi 19, lą 19, fą 19, bn 18, dd 18, rw 18, vp 18, yn 18, fa 17, he 17, ną 17, tl 17, ýk 17, ąf 16, of 16, v# 16, ym 16, zt 16, ąk 15, dp 15, ką 15, kn 15, vÿ 15, wt 15, dw 14, lý 14, nk 14, vz 14, wą 14, wk 14, wm 14, wy 14, yk 14, yu 14, ýc 14, ýl 14, #u 13, di 13, lb 13, us 13, vm 13, yi 13, #e 12, is 12, nÿ 12, tą 12, uÿ 12, vb 12, ým 12, fÿ 11, ub 11, wg 11, ýr 11, ąz 10, dy 10, gą 10, lb 10, rt 10, sl 10, sp 10, yz 10, #f 9, df 9, eo 9, ff 9, kf 9, st 9, sv 9, th 9, wb 9, ýp 9, zf 9, ąl 8, hÿ 8, pÿ 8, tÿ 8, uf 8, ur 8, yd 8, ýv 8, ýd 8, ýi 8, #h 7, ąp 7, bu 7, ge 7, hc 7, hm 7, mf 7, mÿ 7, rl 7, um 7, yb 7, ýn 7, ýf 7, #y 6, ae 6, ąb 6, bc 6, bf 6, dv 6, hl 6, lw 6, r# 6, rk 6, rm 6, ýt 6, ýu 6, gw 5, hr 5, hw 5, mc 5, mv 5, nz 5, rg 5, vd 5, wÿ 5, yr 5, zz 5, aa 4, dm 4, dÿ 4, hi 4, lg 4, my 4, nt 4, se 4, so 4, sw 4, sÿ 4, sz 4, vl 4, vv 4, yv 4, ýw 4, ąs 3, ct 3, ee 3, ef 3, hn 3, ib 3, ii 3, ip 3, kc 3, lt 3, mk 3, ml 3, mr 3, mz 3, nb 3, ng 3, oh 3, pt 3, rr 3, sa 3, sc 3, un 3, vr 3, ya 3, ýo 3, ýb 3, ýp 3, zp 3, bd 2, bf 2, co 2, cÿ 2, fe 2, hą 2, hÿ 2, ig 2, kw 2, nd 2, oe 2, p# 2, pą 2, ps 2, rb 2, rv 2, sg 2, fg 2, ti 2, ul 2, vn 2, wu 2, ye 2, yp 2, ýa 2, ýg 2, ýg 2, ýk 2, ýr 2, ýv 2, zv 2, #i 1, af 1, ao 1, b# 1, bv 1, c# 1, cc 1, cv 1, db 1, dc 1, dg 1, dt 1, ęk 1, f# 1, fv 1, hk 1, hf 1, kk 1, km 1, ky 1, ld 1, lp 1, ls 1, lv 1, lá 1, mg 1, oa 1, oÿ 1, pc 1, pÿ 1, rę 1, sb 1, sd 1, si 1, sk 1, sn 1, ss 1, su 1, sÿ 1, fd 1, tp 1, tv 1, uy 1, uÿ 1, vę 1, vs 1, vt 1, vu 1, vw 1, ws 1, yą 1, yg 1, yo 1, yw 1, ýÿ 1, ýe 1, ýo 1, ýw 1, ýz 1, á# 1, ia 1;

Cresc ie 3334, e# 2297, a# 2260, y# 2081, ni 1460, #w 1448, o# 1439, #p 1352, rz 1308, #t 1252, #f 1135, #z 1129, #n 1048, ze 935, na 934, #i 913, zi 912, i# 885, fz 836, #m 835, #k 803, wi 779, ia 774, #b 763, ch 760, ft 738, dz 736, m# 697, cz 695, za 691, ą# 685, ko 676, ci 675, po 675, ę# 639, wa 632, u# 626, te 624, #c 621, od 602, mi 601, z# 596, fi 590, #o 587, w# 582, #g 571, by 566, ro 566, ey 562, pr 545, ow 543, to 539, #d 534, le 525, an 523, #a 518, em 516, or 502, go 498, ak 495, on 482, of 481, zy 458, en 451, á# 445, #á 443, ar 436, eż 435, ią 431, bo 429, ef 429, ta 429, tr 429, ma 428, ę# 423, ie 407, z# 405, ne 391, af 388, ki 385, zo 385, #r 377, #y 372, ra

366, ży 361, oż 351, ad 343, no 335, oc 330, dy 327, ka 325, #1 323, ło 313, ed 306, ła 305, cz 303, mo 300, ab 299, ob 291, eg 289, kt 279, h# 270, ym 262, wf 260, gd 258, do 257, ec 257, k# 255, t# 253, iu 250, fc 250, że 247, pi 243, aw 242, al 241, im 238, ny 235, da 234, wy 234, yw 233, ai 232, wo 228, ot 221, za 220, in 219, ac 218, yc 204, am 203, yc 202, ep 199, er 197, ku 197, li 196, ol 196, la 195, ek 192, ic 187, ac 186, ba 185, al 183, re 182, ty 182, fa 180, uf 180, el 179, ce 177, eb 177, ok 175, zc 170, ru 167, at 165, ho 161, ac 160, bi 158, io 153, #v 150, ed 149, uc 147, dn 145, op 145, n# 144, iw 143, om 143, fu 143, pa 142, ol 139, br 137, ga 137, iz 136, d# 134, zo 133, dr 129, ry 129, we 127, be 122, ew 122, ag 118, og 117, #z 116, yn 116, gr 115, lu 115, lk 114, tn 114, kr 113, oz 110, pu 110, tk 110, tw 110, if 108, il 107, pl 105, rw 102, ir 101, yf 101, za 101, c# 100, th 100, bu 99, ly 98, mn 98, de 97, fc 96, ui 96, ri 95, zw 95, wl 94, ap 92, fo 90, ha 89, kl 89, #j 88, yt 85, gl 84, yl 84, lk 83, fn 83, et 82, lu 82, fp 82, dl 80, gi 80, l# 79, ez 78, it 78, oi 77, lg 75, rn 75, he 74, dk 73, ic 73, ug 72, az 71, di 71, ei 71, ef 70, na 70, wa 70, yj 70, az 69, ne 69, zk 69, mu 68, zu 68, la 65, zk 65, ad 64, le 64, gn 63, fw 63, ud 63, zb 63, yr 62, ra 60, fl 60, wn 59, zd 59, nf 57, ti 56, uz 56, vz 56, dl 55, je 55, zi 55, zn 55, rc 54, ub 54, j# 53, lo 53, rf 52, wz 52, et 51, s# 51, tl 51, zn 51, ik 50, pn 50, zr 50, ka 49, #s 48, tu 48, kw 47, pl 47, da 46, pl 46, yb 46, za 46, yp 45, ez 44, hc 44, rd 44, ul 43, ij 42, en 40, hy 40, lw 40, ln 39, n# 39, bn 38, dw 37, hl 37, zm 37, at 36, ec 36, lb 36, rk 36, #1 35, me 35, zd 35, wk 34, ek 33, my 33, tc 33, ur 33, wt 33, eb 32, nn 32, ze 32, dm 31, fa 31, hl 30, il 30, fm 30, xi 30, ke 29, fy 29, wr 29, me 28, vr 28, ec 27, kx 27, ut 27, #f 26, be 26, he 26, nk 26, nu 25, fe 25, yl 25, af 24, lk 24, mr 24, r# 24, ta 24, yd 24, zl 24, id 23, vf 23, zf 23, ej 22, hu 22, rm 22, us 22, wu 22, la 21, rp 21, ze 21, zl 21, kc 20, tf 20, we 20, zb 20, ay 19, ge 19, gl 19, oy 19, re 19, uk 19, vp 19, yz 19, dd 18, hn 18, hw 18, ja 18, le 18, te 18, vk 18, zp 18, zu 18, ab 17, ct 17, ml 17, mz 17, zg 17, ag 16, pe 16, pk 16, fl 16, up 16, vc 16, b# 15, ip 15, si 15, yg 15, #e 14, dt 14, ga 14, ib 14, ll 14, lf 14, ml 14, zt 14, zt 14, as 13, du 13, hr 13, km 13, ma 13, nc 13, nb 13, fr 13, au 12, bk 12, co 12, df 12, ez 12, g# 12, gm 12, iz 12, jm 12, tl 12, um 12, zl 12, af 11, de 11, fl 11, fr 11, gu 11, ig 11, l# 11, lc 11, mp 11, rt 11, sa 11, un 11, wc 11, zp 11, ac 10, dc 10, fe 10, hm 10, kn 10, oc 10, on 10, pt 10, vi 10, vw 10, bl 9, bl 9, bf 9, cn 9, fi 9, ln 9, ou 9, p# 9, pe 9, vm 9, wp 9, zl 9, zm 9, jf 8, lw 8, nd 8, os 8, pa 8, so 8, zl 8, #c 7, #h 7, aw 7, ce 7, cu 7, el 7, es 7, ec 7, kk 7, mg 7, py 7, rz 7, st 7, f# 7, zw 7, aa 6, bt 6, ca 6, ci 6, cm 6, ep 6, kf 6, le 6, lt 6, mf 6, rg 6, an 5, ak 5, cw 5, db 5, ha 5, lf 5, pc 5, uw 5, v# 5, wz 5, cy 4, eu 4, hk 4, ia 4, nt 4, nz 4, fd 4, uz 4, vz 4, wd 4, wl 4, ze 4, zr 4, al 3, ar 3, bc 3, dg 3, dp 3, dz 3, eo 3, eg 3, iy 3, kl 3, kz 3, lm 3, mb 3, md 3, nf 3, oe 3, ox 3, rb 3, rl 3, sz 3, fl 3, lz 3, ua 3, uy 3, vt 3, vy 3, ya 3, yi 3, zg 3, #u 2, aj 2, cr 2, fo 2, hb 2, hi 2, hf 2, if 2, jo 2, li 2, nf 2, of 2, rl 2, rr 2, tz 2, vl 2, vo 2, ye 2, aa 1, ah 1, aj 1, ao 1, ax 1, am 1, an 1, ap 1, ba 1, bb 1, bd 1, bm 1,

bż 1, cą 1, cc 1, cj 1, ck 1, cv 1, cw 1, ćb 1, ćl 1, ćf 1, ćz 1, dc 1, ea 1, ęw 1, f# 1, gw 1, gy 1, ht 1, hz 1, ii 1, is 1, ix 1, jd 1, jh 1, jn 1, jr 1, js 1, ke 1, ky 1, kz 1, lż 1, lc 1, ld 1, fl 1, mc 1, mk 1, ng 1, oj 1, ph 1, pj 1, pz 1, sc 1, se 1, sk 1, sp 1, su 1, ue 1, ul 1, uo 1, va 1, vb 1, vd 1, ve 1, vl 1, x# 1, xu 1, yu 1, yz 1, zv 1, źf 1;

Pętla ie 2736, e# 2101, #p 1944, a# 1808, o# 1740, ni 1570, #s 1368, #n 1331, #h 1330, i# 1318, #w 1247, y# 1126, #t 1113, rz 1029, #z 1012, e# 1006, po 1002, ał 991, sz 991, #m 957, #k 933, m# 932, cz 915, wi 902, zy 898, ze 866, dz 836, na 836, #c 811, #d 794, st 729, zi 729, #j 725, ch 722, ia 714, ci 689, ta 685, ie 675, si 673, #o 659, ow 656, z# 645, za 608, #i 604, em 568, #b 557, mi 557, ed 530, la 520, te 503, pr 497, lo 489, ra 487, wa 483, od 481, ki 471, ak 467, ko 457, #r 454, go 451, je 451, an 449, k# 449, li 448, a# 439, es 432, u# 428, ro 425, am 424, ć# 423, pa 420, ja 418, w# 417, ar 413, do 408, ka 400, le 378, to 368, pi 358, eg 352, tr 348, j# 341, #g 327, by 325, er 324, ob 323, ku 321, ty 317, ej 310, ył 309, il 305, wy 301, on 298, ek 295, ba 289, ym 289, ec 287, al 286, #z 285, ys 276, ac 273, aw 271, że 270, bi 265, ad 262, aj 258, oc 248, as 247, ws 246, ub 239, #u 236, ma 233, no 233, at 231, ic 226, #a 224, ot 224, en 223, la 222, #s 221, ne 219, ać 215, os 215, yc 213, ny 211, na 209, my 208, wo 206, h# 203, #l 202, mn 201, kr 200, dy 199, zn 198, mu 197, mo 190, ry 190, n# 188, c# 187, le 183, ów 182, da 181, ja 181, t# 181, om 179, z# 176, us 174, co 171, de 171, ał 169, sk 167, gł 165, tk 164, dn 163, d# 162, ok 159, or 157, az 156, ty 156, oz 156, śc 155, il 154, ap 152, ac 150, is 150, tw 149, dł 148, oś 145, ce 142, ho 140, sł 140, śl 140, ia 138, zu 138, el 137, ha 136, kt 136, yś 136, eś 134, zo 133, ży 133, im 132, #ś 131, lu 130, za 130, zł 130, uc 128, ru 127, oj 126, rę 126, in 125, uc 125, lu 124, ij 123, ór 122, ze 122, br 120, tó 119, uż 116, re 115, cy 114, sp 113, et 111, dr 110, kł 108, yn 108, zę 108, ęd 107, ju 107, ic 106, og 106, op 106, bę 104, zk 103, eb 99, ez 99, mó 99, ol 98, oż 98, yl 98, ga 97, sa 95, zw 95, kn 94, ca 93, ęk 92, lk 90, yt 90, ęc 89, kę 89, aż 87, ep 87, id 87, iw 87, so 87, ag 85, ab 84, um 82, ew 81, se 81, śm 81, gl 80, nt 80, yw 80, be 78, ką 77, ró 77, tu 77, bo 76, za 76, ęt 75, ln 75, eż 74, we 74, pe 73, s# 73, zr 73, is 72, dę 71, n# 71, yp 71, io 70, ur 70, bu 69, eł 67, ik 67, zd 67, gd 66, nę 66, wł 66, wt 66, wr 65, yb 65, aś 64, ađ 64, yk 63, ję 62, py 62, ul 62, śn 61, yj 61, dk 60, rd 58, up 58, wn 58, uj 57, gi 56, rw 56, ut 56, eń 55, zm 55, hc 53, jd 53, oł 53, ój 53, ós 53, ds 52, fi 51, dl 50, ec 50, hw 50, me 50, gn 49, rc 49, wó 49, yc 49, św 48, du 47, wc 47, iu 46, śc 46, #l 45, gr 45, nk 45, #f 44, eł 44, nu 44, su 44, sw 44, tł 44, un 44, bl 43, nó 43, ód 43, rą 43, rt 43, źn 43, hn 42, kl 42, oń 42, pu 42, rn 42, sn 42, yg 42, zl 42, tn 41, ag 40, óc 40, pó 40, ug 40, lo 39, da 38, ln 38, r# 38, ta 38, le 37, uk 37, #h 36, pl 36, sc 36, dw 35, ig 35, jr 35, la 35, nn 35, af 34, fo 34, nc 34, us 34, wk 34, lc 33, rz 32, yd 32, zg 32, cj 31, te 31, uf 31, zó 31, dp 30, ge 30, it 30, oi 30, wa 30, rl 29, ul 29, cu 28, hy 28, ió 28, rm 28, we 28, zd 28, ba 27, hu 27, ps 27, wd 27, wz 27, eś 26,

lk 26, mę 26, pl 26, rs 26, ą 25, dź 25, fe 25, pn 25, di 24, ęż 24, of 24, rk 24, sy 24, hł 23, iz 23, fa 22, gó 22, js 22, mk 22, óg 22, yz 22, zb 22, zt 22, bl 21, uw 21, ań 20, kó 20, nd 20, aż 19, bk 19, cą 19, łó 19, mł 19, mr 19, mt 19, mw 19, ef 18, g# 18, ks 18, ńs 18, ót 18, żo 18, cę 17, śp 17, yr 17, dc 16, jś 16, łb 16, mą 16, ri 16, yż 16, zp 16, ól 15, są 15, tc 15, ż# 15, eż 14, he 14, hi 14, l# 14, nc 14, óz 14, pę 14, sm 14, źc 14, #e 13, ąs 13, lb 13, łą 13, rp 13, ąc 12, bó 12, ck 12, có 12, fl 12, hm 12, jp 12, ls 12, ót 12, tl 12, ąk 11, ep 11, ęs 11, gu 11, jn 11, jo 11, pc 11, pt 11, rg 11, uz 11, zj 11, źn 11, db 10, gą 10, ge 10, hę 10, jc 10, ns 10, oh 10, só 10, wp 10, yy 10, źl 10, żą 10, ąw 9, ęb 9, hl 9, ib 9, jl 9, kw 9, wu 9, zs 9, źm 9, źw 9, źk 9, ai 8, au 8, bc 8, bn 8, cn 8, ęl 8, iz 8, kb 8, kc 8, ld 8, lt 8, mp 8, pk 8, źę 8, aż 7, ąz 7, b# 7, dd 7, dt 7, jg 7, mś 7, ób 7, óz 7, p# 7, rb 7, sf 7, śr 7, #ż 6, dź 6, gw 6, hó 6, in 6, kż 6, łą 6, mg 6, wś 6, dj 5, ęg 5, fn 5, fr 5, ft 5, ii 5, ir 5, ji 5, ke 5, lp 5, ls 5, lt 5, łz 5, ti 5, zu 5, #ó 4, ah 4, eu 4, ęc 4, hr 4, jk 4, ók 4, óz 4, sr 4, źl 4, #y 3, aa 3, ap 3, dm 3, eo 3, fu 3, jb 3, jm 3, kk 3, lg 3, lś 3, mc 3, md 3, ml 3, ms 3, oć 3, ou 3, rc 3, se 3, źc 3, źd 3, źs 3, ay 2, ąb 2, ąś 2, bs 2, cw 2, dó 2, ee 2, ei 2, ip 2, iż 2, jw 2, łc 2, łą 2, nf 2, ńk 2, rl 2, ss 2, ua 2, wb 2, wj 2, wm 2, yo 2, yź 2, zz 2, źó 2, źr 2, bd 1, bj 1, bt 1, cp 1, ęc 1, gm 1, gz 1, hś 1, jź 1, ll 1, lm 1, ló 1, lw 1, mf 1, ng 1, nj 1, nź 1, ńm 1, oo 1, oź 1, pą 1, rś 1, śk 1, śń 1, tp 1, tz 1, uc 1, ui 1, uń 1, wg 1, wl 1, yf 1, yi 1, yń 1, yu 1, źr 1, źb 1;

GG e# 3343, ie 2882, o# 2678, a# 2639, #t 1884, #n 1787, ni 1654, #p 1310, i# 1290, #s 1286, #w 1246, #j 1178, #z 1082, #m 1055, m# 1033, #c 1014, ze 1008, y# 984, na 895, cz 894, #a 873, #d 834, ta 804, z# 798, to 778, ak 777, sz 773, po 742, #b 710, dz 699, le 699, wi 698, ra 693, al 681, za 679, je 675, st 675, ja 670, es 665, ch 644, zy 635, zi 612, te 590, #o 589, rz 588, #k 582, ia 578, si 573, em 571, no 564, am 550, k# 544, ko 539, #i 516, ci 510, ki 505, pr 504, ro 504, do 501, ow 468, w# 452, ma 449, ka 429, mi 429, ac 427, j# 427, wa 406, u# 404, ob 388, #r 373, ty 370, ej 366, an 365, ę# 364, od 362, go 357, ad 353, os 352, as 345, ar 342, mo 336, co 325, er 321, la 321, ec 318, ał 317, c# 314, li 312, by 307, ed 305, da 300, t# 299, en 296, on 294, aj 290, tr 287, aw 279, bi 270, #g 269, my 263, s# 260, ba 259, eg 259, ne 259, pi 256, is 252, ek 249, ęc 245, ym 243, wy 240, az 237, bo 237, oz 236, sp 222, zn 219, at 218, ku 216, or 216, pa 216, ys 215, ie 210, yc 208, łą 200, ic 199, h# 198, ła 196, #l 191, ok 191, re 191, in 185, ju 174, #h 174, mu 171, yl 170, om 169, ol 168, ś# 168, br 166, ło 166, n# 165, ny 164, ez 162, sa 160, we 160, de 159, he 158, że 158, oc 157, ce 155, eb 153, zc 152, il 147, us 145, uz 145, sk 144, #u 142, d# 142, lo 142, ot 140, zo 138, og 136, dn 135, wn 135, ha 134, wo 132, tu 131, ą# 129, ep 129, pe 129, ap 128, be 128, et 128, ry 128, el 127, ws 125, yt 125, tk 124, ew 122, #h 121, l# 121, ż# 121, ga 119, oj 119, im 117, ur 115, lu 114, uj 113, io 112, mn 111, se 111, #e 110, ów 109, sc 109, so 104, op 103, oź 103, ru 101, yb 101, ył 101, ud 100, #ż 99, lk 99, ac 98, dy 98, uc 97, gr 94, id 94,

kt 94, sł 94, ca 93, dr 93, oś 93, #f 91, zk 90, zd 88, ły 86, um 86, eś 85, hy 85, sl 85, sn 85, yw 85, ho 84, rw 84, rt 83, eż 79, nk 79, ęd 78, kr 77, śc 75, ik 74, cj 72, zu 72, bę 70, eh 70, kl 69, rd 69, up 69, zw 68, hc 67, jo 66, ut 66, ił 65, me 64, du 63, iu 63, ks 63, cy 62, nn 61, nt 61, uż 61, yk 61, gi 60, ab 59, ln 59, sy 59, tn 59, wc 59, yn 59, gl 58, is 58, pu 58, tw 58, ag 56, it 56, ró 55, dl 54, r# 54, lu 53, su 53, wr 53, zm 53, fj 52, nc 51, zr 51, wł 50, zę 50, #l 49, ee 49, sm 48, wk 46, iw 45, pl 44, rk 43, ąd 42, dw 42, ub 42, yp 42, fa 41, ns 41, ęc 40, ic 40, ke 40, yg 40, ah 39, aż 39, ia 39, je 39, jn 39, nu 39, py 39, śl 39, uk 39, wd 39, aś 38, iz 38, js 38, yj 38, yś 38, ór 37, zł 37, yc 36, dę 35, ja 35, ly 35, mp 35, gd 34, hi 34, oł 34, un 34, zl 34, hu 33, ug 33, #ś 32, bl 32, gó 32, pł 32, ri 32, sw 32, śn 32, tó 32, ży 32, ąc 31, bu 31, gł 31, rc 31, rr 31, są 31, fo 30, ji 30, lb 30, mó 30, rs 30, tj 30, ul 30, dł 29, ij 29, nd 28, af 27, ai 27, cu 27, dk 27, jd 27, ną 27, oi 27, zt 27, di 26, kc 26, ps 26, zg 26, kę 25, tę 25, ęk 24, kn 24, ój 24, pn 24, pó 24, rn 24, śc 24, wp 24, wu 24, za 24, au 23, bn 23, eł 23, eż 23, kl 23, p# 23, za 23, żn 23, ęt 22, gn 22, la 22, lk 22, rm 22, ka 21, ól 21, re 21, wl 21, fe 20, ng 20, b# 19, he 19, ib 19, kó 19, ńc 19, ds 18, ec 18, eń 18, gu 18, hw 18, jk 18, jm 18, n# 18, sj 18, zj 18, ea 17, oo 17, zp 17, dj 16, g# 16, iz 16, oń 16, ół 16, śm 16, uw 16, wt 16, yd 16, ąc 15, ef 15, eo 15, ge 15, kb 15, wó 15, zb 15, dp 14, jc 14, lc 14, la 14, ób 14, óc 14, św 14, żk 14, cę 13, lę 13, ls 13, łb 13, ti 13, źd 13, źs 13, aż 12, ck 12, gę 12, ir 12, mł 12, sr 12, wz 12, yr 12, yz 12, źl 12, źo 12, bk 11, eu 11, fr 11, hn 11, lw 11, ul 11, ał 10, bc 10, bs 10, ig 10, jś 10, me 10, ne 10, nó 10, ód 10, pk 10, rg 10, só 10, ss 10, zs 10, ań 9, ag 9, có 9, da 9, dc 9, dm 9, ei 9, gw 9, hl 9, jr 9, kw 9, lm 9, lt 9, ln 9, ra 9, we 9, źl 9, cn 8, dó 8, ep 8, hm 8, hó 8, ii 8, ló 8, mk 8, ns 8, óz 8, ta 8, uf 8, wa 8, db 7, #f 7, hl 7, jb 7, jl 7, mm 7, oa 7, oc 7, of 7, ók 7, óz 7, rp 7, sf 7, tl 7, tz 7, us 7, at 6, bó 6, bz 6, ga 6, hh 6, hj 6, hr 6, ip 6, jp 6, jw 6, mb 6, nf 6, nz 6, sr 6, ua 6, #c 5, ao 5, dt 5, eś 5, fu 5, in 5, ll 5, le 5, ms 5, mt 5, np 5, pc 5, uu 5, v# 5, x# 5, yo 5, ź# 5, źc 5, #y 4, as 4, ca 4, cd 4, dd 4, dź 4, ey 4, eb 4, el 4, ez 4, GG 4, jg 4, ky 4, ld 4, lc 4, mś 4, pe 4, tc 4, tl 4, yf 4, yz 4, zó 4, zz 4, źd 4, źn 4, ża 4, #v 3, aż 3, ap 3, az 3, bl 3, cl 3, ct 3, cw 3, dg 3, ex 3, eg 3, el 3, ez 3, ih 3, ió 3, km 3, ló 3, ls 3, mc 3, mh 3, mr 3, ov 3, óg 3, rb 3, rl 3, rl 3, se 3, tm 3, tt 3, yu 3, źm 3, źu 3, #z 2, ax 2, aś 2, ba 2, cc 2, cm 2, cw 2, cn 2, dź 2, es 2, fk 2, fn 2, gz 2, jj 2, jt 2, kp 2, kś 2, lp 2, lź 2, li 2, mf 2, mg 2, ml 2, nx 2, nk 2, ox 2, ós 2, ót 2, pa 2, pt 2, rf 2, rs 2, śp 2, th 2, tp 2, tv 2, ve 2, wf 2, wg 2, wj 2, xa 2, xp 2, ye 2, źc 2, źb 2, #ó 1, #q 1, av 1, ay 1, ab 1, an 1, bh 1, bm 1, bt 1, cp 1, cs 1, cś 1, cz 1, cc 1, cm 1, df 1, dś 1, ev 1, ei 1, ej 1, em 1, en 1, eo 1, ff 1, fl 1, fm 1, fy 1, fz 1, gh 1, gm 1, ha 1, hd 1, ht 1, hx 1, if 1, ix 1, iz 1, jh 1, jz 1, kj 1, kk 1, kz 1, kź 1, lz 1, ld 1, ma 1, mj 1, nb 1, nh 1, nj 1, nw 1, nć 1, oe 1, oh 1, ou 1, oy 1, oź 1, ó# 1, óa 1, óz 1, pm 1, pp 1, pw 1, pz 1, qr 1, sb 1, sc 1, sd 1, sb 1, sg 1, śl 1, śs 1, tf 1, tg 1, ts 1, uh 1, ui 1, ux 1, va 1, vi 1, vo 1, wm 1, ww 1, xt 1, xu 1, xy 1, ya 1, yi 1, yź 1, źt 1, źc 1, źr 1, źz 1.

7.3. Dyspersja diad znakowych

Średnia częstość występowania diad liter (F_2) jest w poszczególnych próbach tekstowych różna – *Gaszt*: $F_2 = 128$, *Cresc*: $F_2 = 150$, *Pętla*: $F_2 = 153$, *GG*: $F_2 = 131$. W analizie porównawczej statystycznego rozkładu dwuelementowych połączeń liter zastosowano segmentację frekwencyjną jednostek użytą przez Kuraszkiewicza do badań nad statystyką leksemów²³. Wedle tego schematu diady dzielą się na następujące grupy frekwencyjne:

- diady o najwyższej częstości – 1% wszystkich diad tekstu,
- diady bardzo częste – 4% wszystkich diad tekstu,
- diady niezbyt częste, ale o częstości wyższej niż przeciętna F_2 ,
- diady o częstości przeciętnej, czyli równej F_2 ,
- diady rzadkie, o częstości poniżej F_2 , ale większej od 1,
- diady bardzo rzadkie (wyjątkowe), o częstości równej 1.

Dyspersję diad literowych *Gaszt*, *Cresc*, *Pętli* i *GG* pod względem ich frekwencji tekstowej w wyszczególnionych grupach ilustruje tabela 3.

Na poziomie repertuaru diad literowych różnice między omawianymi próbami tekstowymi dotyczą trzech pierwszych i dwóch ostatnich grup. W przypadku grup 1. i 2. odmiennymi wskazaniami wyraźnie odznacza się *GG*: grupę 1. stanowi w tej próbie zbiór 7 diad (w pozostałych fragmentach tekstów – 6 diad), grupę 2. zaś – 30 diad (w grupie tej liczba diad przyrasta co prawda chronologicznie: *Gaszt* – 23 diady, *Cresc* – 25 diad, *Pętla* – 26 diad, jednak największy wzrost ich notowań zaznaczył się w *GG* – 4 diady więcej niż w *Pętli*). Niemniej nie należy tych wskazań interpretować jako świadczących o najmniejszym zróżnicowaniu diad znakowych w *GG*, gdyż są one bezpośrednio zależne od liczby diad poświadczonych w danej próbie – w przypadku *GG* najwyższej (746 // *Gaszt* 567 // *Cresc* 635 // *Pętla* 648). W zakresie wartości obliczonych dla grup 3., 5. i 6. analizowane teksty przyporządkować można dwóm zbiorom. Pierwszy z nich, do którego zaliczyć należy *Gaszt*, *Cresc* i *GG*, składa się z prób poświadczających chronologiczny (choć nie ciągły, ponieważ nie dokumentuje go *Pętla*, zob. niżej) wzrost wskazań dla diad z grup 5. i 6. (*Gaszt* 63,1% i 10,8% // *Cresc* 64,1% i 11,3% // *GG* 65,3% i 12,5%), któremu towarzyszy chronologiczny spadek odsetka liczby diad w grupie 3. (*Gaszt* 20,8% // *Cresc* 19,5% // *GG* 17,2%). Wskazania te oznaczają, że obrazowane przez te próby przekształcanie systemu graficznego odznacza się przyrostem diad o niskich i najniższych frekwencjach tekstowych, czyli zwiększaniem jego zróżnicowania. Odmiennie wskazania w tych grupach poświadczą *Pętla*; w tekście tym zwraca uwagę wysoki odsetek diad rzadkich (69,3% // *Gaszt*

²³ W. Kuraszkiewicz, op.cit., s. 696–708.

63,1% // *Cresc* 64,1% // *GG* 65,3%) i bardzo niski (od wskazań obliczonych dla pozostałych prób niższy średnio o połowę) wskaźnik obliczony dla diad bardzo rzadkich (5,7% // *Gaszt* 10,8% // *Cresc* 11,3% // *GG* 12,5%). W próbie tej zauważalny jest zatem spadek poświadczeń diad o frekwencji równej 1 na korzyść grupy połączeń rzadkich (o frekwencji wyższej od 1, ale niższej niż średnia arytmetyczna częstości występowania diad w tekście) – tym samym wskazania te świadczą o zaznaczającej się we współczesnej grafii standardowej tendencji do zmniejszania różnorodności systemu na poziomie dwuelementowych połączeń literowych.

Tabela 3. Dyspersja diad znakowych

	Gr. fr.	Diady				Użycia		Stosunek grup 1., 2., 3., do 4., 5., 6.			
		ranga	frekwencja	liczba (L_D)	$\%L_D$	liczba (U_D)	$\%U_D$	diady		użycia	
								liczba (L_D)	$\%L_D$	liczba (U_D)	$\%U_D$
<i>Gaszt</i>	1.	1–6	3 042–1 336	6	1,0	11 353	15,6	147	26	61 878	85
	2.	7–28	1 181–575	23	4,0	17 423	23,9				
	3.	29–127	566–129	118	20,8	33 102	45,5				
	4.	128	128	1	0,2	128	0,2	420	74	10 943	15
	5.	129–223	126–2	358	63,1	10 754	14,8				
	6.	224	1	61	10,8	61	0,1				
<i>Cresc</i>	1.	1–6	3 334–1 448	6	1,0	12 880	13,6	155	24	79 473	84
	2.	7–31	1 439–685	25	4,0	23 034	24,3				
	3.	32–141	676–153	124	19,5	43 559	45,9				
	4.	142	150	1	0,2	150	0,2	480	76	15 453	16
	5.	143–256	149–2	407	64,1	15 231	16,0				
	6.	257	1	72	11,3	72	0,1				
<i>Pętla</i>	1.	1–6	2 736–1 570	6	1,0	11 899	12,0	162	25	80 949	81
	2.	7–29	1 368–729	26	4,0	26 002	26,2				
	3.	30–146	725–154	130	20,1	43 048	43,3				
	4.	–	–	–	–	–	–	486	75	18 442	19
	5.	147–272	152–2	449	69,3	18 405	18,5				
	6.	273	1	37	5,7	37	0,04				
<i>GG</i>	1.	1–7	3 343–1 654	7	1,0	16 867	17,2	165	22	82 624	84
	2.	8–37	1 310–675	30	4,0	26 863	27,5				
	3.	38–145	670–132	128	17,2	38 894	39,7				
	4.	146	131	1	0,1	131	0,1	581	78	15 228	16
	5.	147–254	129–2	487	65,3	15 004	15,3				
	6.	255	1	93	12,5	93	0,1				

Dyspersja użyć diad literowych pokazuje, że próbą tekstową w największym stopniu nasyconą połączeniami znakowymi o najwyższych i bardzo wysokich frekwencjach jest *GG* (grupa 1.: 17,2% // *Gaszt* 15,6% // *Cresc* 13,6% // *Pętla* 12,0%; grupa 2.: 27,5% // *Gaszt* 23,9% // *Cresc* 24,3% // *Pętla* 26,2%); w odróżnieniu od pozostałych tekstów *GG* poświadcza za to najmniejszy pułap wystąpień diad z grupy 3. (39,7% // *Gaszt* 45,5% // *Cresc* 45,9% // *Pętla* 43,3%) – oznacza to, że w grupie połączeń częstych (pierwsze trzy grupy frekwencyjne) rozkład użyć diad w *GG* jest najbardziej nierównomierny. *GG* poświadcza też najniższy pułap użyć dwuelementowych połączeń literowych w grupie 4. i dość niski odsetek wystąpień diad w grupie 5., co oznacza, że strukturę reprezentowanego przez tę próbę tekstową systemu graficznego w większym stopniu niż w przypadku pozostałych analizowanych w niniejszym artykule struktur graficznych tworzy niewielka grupa diad o wysokich frekwencjach tekstowych. Pod względem użyć diad o frekwencji równej 1 *GG* nie wykazuje odstępstw od większości analizowanych prob (prócz *Pętli*), utrzymując się na poziomie 0,1%.

Wśród pozostałych prób tekstowych wyróżnia się też *Pętla*. Tekst ten charakteryzuje najniższy odsetek wystąpień diad należących do dwóch najbardziej skrajnych grup frekwencyjnych (grupa 1.: 12,0% // *Gaszt* 15,6% // *Cresc* 13,6% // *GG* 17,2%; grupa 6.: 0,04% // *Gaszt*, *Cresc*, *GG* 0,1%). Zaznaczyć jednak należy, że w obu przypadkach spadek wartości notowań diad rekompensowany jest wzrostem (w porównaniu z tekstami chronologicznie wcześniejszymi) wystąpień połączeń literowych z grup sąsiadujących, czyli 2. i 5. (grupa 2.: 26,2% // *Gaszt* 23,9% // *Cresc* 24,3% // *GG* 27,5%; grupa 5.: 18,5% // *Gaszt* 14,8% // *Cresc* 16,0% // *GG* 15,3%). Zjawisko w obrębie relacji między spadkiem użyć diad z 1. grupy a wzrostem wystąpień połączeń z 2. grupy frekwencyjnej zaznaczyło się też w innych próbach pochodzących z okresów wcześniejszych; poświadczają one także chronologiczny wzrost wystąpień diad z grupy 5. – w żadnej z nich jednak zmianom tym nie towarzyszy tak diametralny spadek notowań połączeń o frekwencji równej 1. Dyspersja porównawcza użyć diad pozwala zatem scharakteryzować współczesną grafikę standardową jako system tworzony w większej mierze niż pozostałe analizowane w niniejszym artykule systemy graficzne przez diady bardzo częste i rzadkie – choć w zakresie użyć diad należących do grupy połączeń rzadkich system ten charakteryzuje się najwyższą wartością odsetkową.

Ujawniona dysproporcja między wskazaniem dla liczby i użyć diad wyjątkowych w *GG* (grupa 6.: bardzo wysoki odsetek liczby diad // przeciętny odsetek użyć diad), uwidaczniająca się również w zbiorczym zestawieniu wartości właściwych dla grup frekwencyjnych diad częstych i rzadkich (najwyższy procent liczby diad rzadkich // największy odsetek użyć diad częstych), świadczy

o wysokich frekwencjach diad należących do pozostałych, zwłaszcza pierwszych dwóch grup frekwencyjnych w tej próbie tekstowej. Na poziomie analizy diad *GG* wykazuje się zatem tą samą właściwością, która ujawniła się w tej próbie tekstowej w zakresie wskazań cechujących rozkład liter: z jednej strony odznacza się ona najwyższymi ze wszystkich przebadanych prób wskaźnikami obliczonymi dla liter częstych i diad bardzo częstych i częstych, z drugiej zaś – charakteryzuje się wysokimi wartościami uzyskanymi dla liter i diad literowych bardzo rzadkich. Uznać zatem należy, że – w odróżnieniu od właściwości statystycznych ujawnianych przez rozkład liter i ich diad we współczesnym znormalizowanym systemie graficznym, bliższych wartości średniej – innowacyjna grafia mediów elektronicznych wykazuje tendencje do preferowania znaków i diad znakowych należących do skrajnych grup frekwencyjnych. Tę opozycję wskazań charakteryzujących *Pętłę* i *GG* trudno zinterpretować wedle wykorzystywanego w niniejszej pracy schematu: nie sposób bowiem przyznać systemowi graficznemu jednej z tych prób wyższego stopnia różnorodności; pod tym względem uznać należy te systemy za równoważne.

7.4. Entropie diad znakowych

Wartości parametrów teoriiinformacyjnych obliczone dla diad znakowych poświadczonych w analizowanych próbkach tekstów prezentuje tabela 4.:

Tabela 4. Entropie diad literowych (pierwszego rzędu): bezwzględna, względna; redundancja względna diad

	<i>Gaszt</i>	<i>Cresc</i>	<i>Pętla</i>	<i>GG</i>
H_2	3,26	3,34	3,40	3,39
H_w^2	0,66	0,66	0,67	0,66
R_2	0,34 (34%)	0,34 (34%)	0,33 (33%)	0,34 (34%)

Z wartości entropii rzędu drugiego (H_2) wynika, że systemem graficznym o najwyższej pojemności informacyjnej jest system reprezentowany przez *Pętłę* (3,40 bita na literę). Nieco mniejszą wartość wskaźnik ten przyjmuje dla *GG* (3,39 bita na literę), zaś w tekstach XVI-wiecznych wartość tego parametru teoriiinformacyjnego jest zbliżona, choć nieznacznie przeważa na korzyść *Cresc* (*Gaszt*: 3,26 bita na literę; *Cresc*: 3,34 bita na literę).

W zakresie wartości entropii względnej rzędu drugiego analizowane teksty tworzą hierarchię zgodną z rozkładem wielkości entropii względnej rzędu pierwszego (H_w^1). Najwyższą wartość entropia względna diad (H_w^2) przyjmuje w *Pętli* (0,67; redundancja R_2 na poziomie 33%), tworząc tym samym opozycję w stosunku do mniejszych i jednocześnie bardzo zbliżonych (iden-

tycznych po zaokrągleniu) wartości tego wskaźnika obliczonych dla tekstów XVI-wiecznych i dla tekstów reprezentujących grafikę mediów elektronicznych (*Gaszt*, *Cresc*, *GG*: $H_w^2 - 0,66$; $R_2 - 34\%$)²⁴. Świadczy to o tylko nieco większym stopniu uzależnienia występowania znaku od jego poprzednika we współczesnej grafii standardowej niż w pozostałych badanych systemach pisownianych (w których, notabene, stopień takiego związania jest nieomal identyczny), co oznacza, że i na tym poziomie to współczesny znormalizowany system graficzny odznacza się największą, choć zbliżoną do pozostałych systemów pisownianych, efektywnością informacyjną.

8. Wnioski

Na podstawie dokonanych obliczeń sformułować można następujące wnioski:

1. Systemem graficznym o najmniejszym stopniu dyferencjacji rozkładu znaków (na poziomie zarówno liter, jak i ich diad) jest współczesny znormalizowany (standardowy) system pisowniany. Pod tym względem dzieje polskiego systemu graficznego do momentu wykrystalizowania się współczesnego systemu standardowego wykazują tendencję porządkującą (zaznaczającą się już w przeanalizowanych systemach XVI-wiecznych) polegającą na redukowaniu znaków i dwuelementowych połączeń znakowych o frekwencji równej 1.

2. Struktura innowacyjnej grafii mediów elektronicznych w większym stopniu niż pozostałe badane systemy graficzne opiera się na literach i diadach należących do skrajnych grup frekwencyjnych, czyli na literach i diadach literowych najczęstszych i najrzadszych (o frekwencji równej 1). Pod tym względem stanowi ona przeciwieństwo struktury statystycznej współczesnej grafii standardowej, która współtworzona jest głównie przez litery i diady należące do grup frekwencyjnych bliższych wartości średniej. Oznacza to, że innowacyjne rozwiązania stosowane w grafii mediów elektronicznych nie naruszają statystycznej proporcji między literami i ich diadami właściwymi dla współczesnej grafii standardowej.

3. Pod względem pojemności informacyjnej wśród analizowanych systemów graficznych nieznacznie wyróżnia się współczesna pisownia znormalizowana, odznaczając się najwyższą wartością entropii względnej najwyższego z obliczonych w niniejszej analizie rzędów – rzędu drugiego. Pozostałe systemy charakteryzują się nieomal identycznymi wskazaniem stopnia swobody wyboru liter (po zaokrągleniu do ułamka setnego są to wartości identyczne). Oznacza to, że reprezentowane przez analizowane próby tekstowe systemy graficzne, w analizach jakościowych traktowane jako bardzo różne, pod wzglę-

²⁴ Przed zaokrągleniem – H_w^2 : *Gaszt* 0,657663434 // *Cresc* 0,655680408 // *GG* 0,656391757; wskaźnik ten jest zatem najniższy w *Cresc*.

dem redundancji pozostają tożsame (systemy z pierwszej połowy XVI wieku i grafia mediów elektronicznych – redundancja 33%) lub bardzo sobie bliskie (współczesna grafia standardowa a pozostałe badane systemy: 34% // 33%). Z przeprowadzonej analizy wynika, że pod względem teoriiinformacyjnym przekształcenia systemu graficznego (w analizowanym zakresie chronologicznym) nie polegają ani na postępie (ewolucji), ani na degradacji.

Marek Osiewicz, Krzysztof Skibski

The statistical structure of the Polish graphic system.

A diachronic and comparative analysis

This article presents a statistical and comparative analysis of four spelling conventions that represent different stages in the development of the Polish graphic system: the graphic system of a late-medieval manuscript (hand-written text), the standard spelling convention typical for the first half of the sixteenth century, the accepted and standard modern spelling of the first half of the twentieth century and the innovative set of graphic features used in electronic media. The characteristics of the statistical parameters encompasses dispersion and entropy in the first and the second row of letters, as well as in two-element sets (dyads). The analysis proves that: 1) inasmuch as the degree of differentiation of the distribution of signs, the history of Polish spelling convention prior to the solidification of the modern standard practice (accepted standard system) manifested a self-organizing tendency that was based on a reduction of letter signs and two-element letter combinations (ligatures) with the frequency of 1; 2) innovative solutions used in the set of graphic features characteristic for electronic media do not violate the statistical proportion between letters and their dyads operative and specific for modern standard graphic system 3) in respect to theory and information, the transformations of the graphic substance (graphic system) (within the analysed chronological timeframe) depended on neither progress (evolution) nor degradation.

KEY WORDS: history of the Polish language, modern standard Polish, graphic system, quantitative linguistics, comparative analysis, theory of information.

dr Marek Osiewicz, Zakład Historii Języka Polskiego, Instytut Filologii Polskiej Uniwersytetu im. Adama Mickiewicza w Poznaniu; zainteresowania naukowe: polszczyzna pierwszej połowy XVI wieku, fonetyka, fonologia i morfologia historyczna, metodologia badań historycznojęzykowych.

dr Krzysztof Skibski, Zakład Frazeologii i Kultury Języka Polskiego, Instytut Filologii Polskiej Uniwersytetu im. Adama Mickiewicza w Poznaniu; zainteresowania naukowe: język współczesnych tekstów literackich, składnia poetycka, tendencje rozwojowe polszczyzny, metodologia badań językoznawczych.