„LingVaria”
Rok IV (2009) nr 2 (8)

Agnieszka Kułacka
King's College
University of London
London

PROCEDURA WERYFIKACJI PRAWA KRYŁOWA

Wstęp

Prawo Kryłowa dotyczy polisemii i opisuje zależność między liczbą *y* leksemów w słowniku danego języka posiadających *x* znaczeń a liczbą *x*. Pierwszy zauważył tę prawidłowość George K. Zipf w 1949 roku, badając jednojęzyczny słownik angielski. Dostrzegł on, że wraz ze wzrostem liczby znaczeń, *x*, liczba leksemów mających *x* znaczeń spada i wyraził to funkcją:

$$y = \frac{C}{x^2}\ , (1)$$

gdzie *C* jest pewną stałą(Hammerl, Sambor 1993; Sambor 1972). Ten model był później krytykowany między innymi przez Annę Wierzbicką (Sambor 1972: 64).

Badania nad prawem podjął w 1967 roku Ferenc Papp, analizując 60 tysięcy leksemów słownika języka węgierskiego. Dostrzegł, że dane empiryczne można ująć pewną funkcją ściśle malejącą[1]. Największą grupę stanowiły leksemy jednoznaczeniowe, dużo mniejszą leksemy o dwóch znaczeniach, a znikomą część leksemy wieloznaczeniowe. Papp dla swoich danych przyjął poniższy model teoretyczny:

$$y = \frac{W}{2^x}\ , (2)$$

gdzie *W* jest liczbą leksemów w słowniku jednojęzycznym. Trudno tutaj mówić o dobrym dopasowaniu danych empirycznych i teoretycznych tego modelu, gdyż F. Papp nie podał w swojej pracy żadnych danych (Sambor 1972: 75).

W 1982 roku Jurij U. Kryłow zebrał dane z dwóch rosyjskich słowników jednojęzycznych. Dostrzegł, że dane te przedstawiają podobną strukturę statystyczną,

[1] Mówimy, że funkcja f jest ściśle malejąca w przedziale (a,b), jeśli dla każdego x,y należących do tego przedziału i spełniających nierówność x < y, mamy f(x) > f(y).

z czego wyciągnął wniosek, że różnice liczbowe powstały tylko na skutek niejednoznacznych kryteriów wydzielania znaczenia wyrazów, a badana przez niego tendencja może pretendować do rangi prawa statystycznego. Kryłow zaproponował inny model teoretyczny:

$$p_x = \frac{1}{2^x} \cdot \frac{1}{1 - p_1},$$

gdzie p_x jest prawdopodobieństwem wylosowania ze słownika jednojęzycznego leksemu o x znaczeniach, a p_1 jest prawdopodobieństwem wylosowania leksemu o jednym znaczeniu (Krylov 1982: 250). Model ten w późniejszych latach okazał się niezadowalający, gdyż zastosowanie testów empirycznych pozwoliło stwierdzić, że rozkłady empiryczny i teoretyczny nie są zbieżne (Hammerl, Sambor 1993: 124).

W 1990 roku Jadwiga Sambor opublikowała wyniki badań nad prawem Kryłowa, jakie przeprowadziła na słownikach jednojęzycznych języków polskiego, rosyjskiego i angielskiego, losując próby o określonej liczbie leksemów. Badaczka porównała dane empiryczne uzyskane przez siebie i przez Kryłowa i dostrzegła zbieżność między tymi rozkładami. Zaobserwowane różnice odniosła do metodologii wydzielania znaczeń oraz do charakterystyki danego języka, na przykład słownik języka angielskiego ma więcej leksemów wieloznacznych niż słowniki języka polskiego i rosyjskiego (Hammerl, Sambor 1993: 120–123).

Badania nad prawem Kryłowa trwają nadal. Gabriel Altmann wyjaśnia konieczność jego zachodzenia siłami Zipfa leżącymi u podstaw zachowania ludzkiego, przestrzegającego zasady najmniejszego wysiłku: nadawca i odbiorca usiłują zapewnić sobie minimum wysiłku przy nadawaniu i rozumieniu tekstu. Rezultatem działania tych sił są procesy unifikacji i dywersyfikacji w języku (por. Sambor 1989; Altmann 2005). Zmodyfikowany model teoretyczny opisujący prawo Kryłowa znajduje się w pracy Rolfa Hammerla (1991), do której autorzy pracy *O statystycznych prawach językowych* odsyłają (Hammerl, Sambor 1993: 125). Inne modele teoretyczne znajdują się w opracowaniu Gezji Wimmera i Gabriela Altmanna (2005), jednakże żaden z nich nie został w pełni[2] zweryfikowany.

W niniejszym artykule podam moje spostrzeżenia dotyczące sposobu weryfikacji prawa Kryłowa, a więc odpowiem na pytania, czy można losowo dobierać próbę oraz czy otrzymany rozkład częstości względnych jest zbieżny z pewnym rozkładem prawdopodobieństw. Jak widzimy, badacze stosują różne procedury: jedni losowo wybierają próby, inni badają cały słownik jednojęzyczny. Uznałam więc, że należy doprecyzować sposób weryfikacji prawa Kryłowa.

Funkcja opisująca prawo Kryłowa

Jak już wiemy ze wstępu do tego artykułu, nie istnieje zadowalający model funkcyjny, który opisywałby prawo Kryłowa. Dostrzec można jednak pewne własności tej funkcji, nazwijmy ją f, która liczbie znaczeń leksemów x przypisuje ich częstości y:

[2] Badania wykonano na małych próbach, ograniczających się zwykle do jednego języka.

$$y = f(x) . (4)$$

Własności funkcji *f*:
(1) Jest to funkcja dyskretna. Zmienna niezależna *x* przyjmuje wartości 1, 2, 3, …
(2) Jest to funkcja ściśle malejąca dla 99,5% badanych leksemów o najmniejszych liczbach znaczeń. Przyjrzyjmy się danym w tabeli 1:

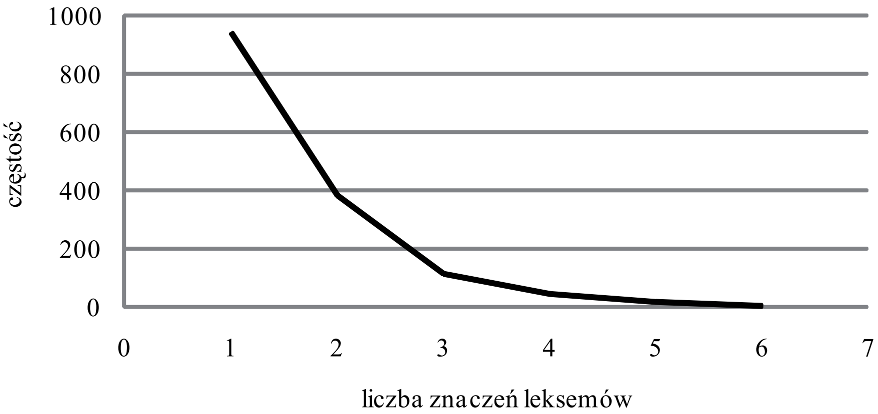
Tabela 1. Częstości leksemów x-znaczeniowych zaczynających się na literę u

x	Częstość	Częstość względna, R(x)
1	941	0,6195
2	385	0,2535
3	113	0,0744
4	46	0,0303
5	20	0,0132
6	7	0,0046
7	1	–
8	4	–
9	1	–
10	1	–
Suma	1519	$\approx$ 0,995

99,5% z liczby 1519 to z dokładnością do liczby całkowitej 1511, czyli liczba leksemów o liczbie znaczeń mniejszej niż 7. Częstości dla tych leksemów ściśle maleją.
(3) Teoretyczny model ciągły przybliżający funkcję *f* będzie wypukły, a to znaczy, że łuk wykresu tego modelu łączący dwa dowolne punkty P, Q tego wykresu leży poniżej lub na cięciwie PQ. Przyjrzyjmy się wykresowi funkcji *f*, sporządzonemu dla danych zawartych w tabeli 1. Punkty na wykresie zostały połączone odcinkami, które są jednocześnie cięciwami.
(4) Asymptotą pionową dla teoretycznego modelu ciągłego przybliżającego funkcję *f* będzie oś rzędnych. Zmienna *y*, przy wartościach zmiennej *x* zbliżających się do 0, przyjmuje wartości coraz większe. W tabeli 1 oraz na rysunku 1 wartość funkcji *f* dla *x = 1* jest największa. Z własności (3) wynika, że w potencjalnym teoretycznym modelu, przy zbliżaniu się wartości *x* do zera wartości funkcji *f* będą dążyć do nieskończoności.
(5) Asymptotą poziomą dla teoretycznego modelu ciągłego przybliżającego funkcję *f* będzie oś odciętych. Zmienna *y*, przy wartościach zmiennej *x* zbliżających się do nieskończoności, przyjmuje wartości coraz mniejsze. Przyjrzyjmy się wartościom funkcji *f* w tabeli 1: dla dużych wartości *x* wartości funkcji maleją do zera.

By móc porównać wartości częstości o danej liczbie znaczeń leksemów dla prób o różnych liczbach leksemów, można obie strony równania (4) opisującego funkcję *f* podzielić przez liczbę badanych leksemów i otrzymać wzór opisujący zależność między względnymi częstościami o liczbą znaczeń leksemów:

Rysunek 1. Wykres funkcji ***f***



$$R(x) = \frac{f(x)}{W}, (5)$$

gdzie *W* jest liczbą badanych leksemów. Dla przykładu w tabeli 1 mamy:

$$R(1) = \frac{941}{1519} \approx 0{,}6195$$

W moim badaniu porównywać będę częstości względne.

Wybór próby

Korzystając z internetowej wersji *Słownika języka polskiego PWN*, wybrałam kilkanaście liter, dla których obliczyłam liczbę leksemów mających $x = 1, 2, 3, \ldots$ znaczeń. Cała badana przeze mnie próba liczy 20 413 leksemów i dobrana została metodą przypadkową (ang. *convenience sampling*), czyli dobór próby wykonano na podstawie subiektywnej choć nie tendencyjnej decyzji badacza. Analizowane były leksemy zaczynające się na następujące litery: *a*, *b*, *c*, *ć*, *d*, *f*, *h*, *ó*, *q*, *ś*, *u*, *x*, *y*, *z*, *ź*, *ż*. Obliczyłam i porównywałam względne częstości dla 99,5% leksemów o największych częstościach, stanowiące jednocześnie zbiór leksemów o najmniejszej liczbie znaczeń. Pozostałe 0,5% uznałam za leksemy występujące rzadko, powodujące zakłócenie ścisłej monotoniczności funkcji. Stosując inny sposób analizy wyników, wykorzystujący współczynnik korelacji rangowej Spearmana, który przedstawiałam w *Podsumowaniu*, brałam pod uwagę całość próby i uzyskałam ten sam wniosek, a mianowicie, że prawo Kryłowa zachodzi. Poniższy sposób analizy będzie w przyszłości prowadził do znalezienia odpowiedniego modelu matematycznego, a więc pewnej funkcji służącej do opisu tego prawa, oraz wysunięcia wniosków dotyczących występujących we wzorze funkcji współczynników. W tabeli 2 umieszczone są zebrane dane wraz z obliczonymi częstościami względnymi.

Tabela 2. Częstości leksemów *x*-znaczeniowych dla całej próby

x	Częstość	Częstość względna, R(x)
1	14089	0,6902
2	4004	0,1961
3	1315	0,0644
4	516	0,0253
5	236	0,0116
6	120	0,0059
7	50	0,0024
8	28	–
9	17	–
10	10	–
11	9	–
12	4	–
13	5	–
14	2	–
15	4	–
17	1	–
18	2	–
27	1	–
Suma	20413	≈ 0,996

Następnie z otrzymanej populacji wybrałam różne populacje o mniejszej, ale zbliżonej liczbie leksemów, około 17 268 leksemów. Miało to na celu postawienie dotyczącej sposobu wyboru próby hipotezy, według której przy zachowaniu zbliżonej liczby leksemów, ale po zmianie jakości próby uzyskuje się bardzo zbliżone wartości częstości względnych. Dzięki zastosowanemu niżej aparatowi statystycznemu, wykazałam, że zawartość jakościowa próby nie wpływa na wartości częstości względnych. W dalszej części artykułu przyjrzymy się, jaka jest wystarczająca szacunkowa wielkość próby.

Dane dotyczące częstości leksemów *x*-znaczeniowych dla różnych prób umieszczone są w tabelach 3–8.

Tabela 3. Częstości leksemów *x*-znaczeniowych dla próby po usunięciu leksemów zaczynających się na literę *d*

x	Częstość	Częstość względna, *R(x)*
1	11917	0,6902
2	3393	0,1965
3	1115	0,0646
4	442	0,0256
5	182	0,0105
6	105	0,0061
7	41	0,0024
Suma	wszystkich leksemów: 17266	$\approx$ 0,996

Tabela 4. Częstości leksemów *x*-znaczeniowych dla próby po usunięciu leksemów zaczynających się na literę *a*, *ć*, *ś*, *y*

x	Częstość	Częstość względna, *R(x)*
1	11776	0,6820
2	3464	0,2006
3	1150	0,0666
4	438	0,0254
5	218	0,0126
6	105	0,0061
7	44	0,0025
Suma	wszystkich leksemów: 17268	$\approx$ 0,996

Tabela 5. Częstości leksemów *x*-znaczeniowych dla próby po usunięciu leksemów zaczynających się na literę *b*, *ć*, *ż*

x	Częstość	Częstość względna, *R(x)*
1	11928	0,6907
2	3361	0,1946
3	1140	0,0660
4	458	0,0265
5	207	0,0120
6	92	0,0053
7	45	0,0026
Suma	wszystkich leksemów: 17269	$\approx$ 0,998

Tabela 6. Częstości leksemów *x*-znaczeniowych dla próby po usunięciu leksemów zaczynających się na literę *ć*, *h*, *ó*, *q*, *u*, *x*, *ź*, *ż*

x	Częstość	Częstość względna, *R(x)*
1	11913	0,6898
2	3364	0,1948
3	1114	0,0645
4	437	0,0253
5	200	0,0116
6	103	0,0060
7	48	0,0028
Suma	wszystkich leksemów: 17271	≈ 0,995

Tabela 7. Częstości leksemów *x*-znaczeniowych dla próby po usunięciu leksemów zaczynających się na literę *ć*, *f*, *h*, *q*, *ź*, *ż*

x	Częstość	Częstość względna, *R(x)*
1	11712	0,6784
2	3504	0,2030
3	1154	0,0668
4	459	0,0266
5	203	0,0118
6	110	0,0064
7	48	0,0028
Suma	wszystkich leksemów: 17265	≈ 0,996

Tabela 8. Częstości leksemów *x*-znaczeniowych dla próby po usunięciu leksemów zaczynających się na literę *ć*, *f*, *q*, *u*, *ź*

x	Częstość	Częstość względna, *R(x)*
1	11937	0,6912
2	3356	0,1943
3	1117	0,0647
4	432	0,0250
5	198	0,0115
6	112	0,0065
7	48	0,0028
Suma	wszystkich leksemów: 17270	≈ 0,996

Obliczę teraz 95% przedziałów ufności[3] dla średnich częstości względnych leksemów o danej liczbie znaczeń, zgodnie ze wzorem:

[3] Dziewięćdziesięciopięcioprocentowy przedział ufności oznacza, że mamy 95% pewność, iż wartość średniej badanych populacji prób o danej liczbie leksemów znajdzie się w danym przedziale.

$$\overline{R(x)} - t_{0,975}\frac{s}{\sqrt{N-1}} < \mu(x) < \overline{R(x)} + t_{0,975}\frac{s}{\sqrt{N-1}}, \ (6)$$

gdzie $\overline{R(x)}$ jest średnią próby, *N* – liczbą prób, *s* – odchyleniem standardowym próby, μ jest średnią wszystkich prób liczących około 17 268 leksemów w słowniku. Występująca we wzorze $t_{0,975}$ to wartość *t* w rozkładzie Studenta (właściwego dla małych prób), dla której 2,5% pola znajdującego się pod wykresem funkcji gęstości leży w prawostronnym ogonie rozkładu.

W przypadku powyższego badania $N = 6$ (dane zawarte w tabelach 3–8), a odczytane z tablic statystycznych $t_{0,975} = 2,57$ przy $\nu = N - 1$, tj. $\nu = 5$ stopniach swobody. W tabeli 9 umieszczone są obliczone przedziały ufności.

Tabela 9. Przedziały ufności dla średnich częstości względnych leksemów o danej liczbie znaczeń

x	$\overline{R(x)}$	s	Przedział ufności
1	0,6871	0,0050	0,6816 < μ(1) < 0,6928
2	0,1973	0,0033	0,1935 < μ(2) < 0,2011
3	0,0655	0,0010	0,0644 < μ(3) < 0,0666
4	0,0257	0,0006	0,0250 < μ(4) < 00264
5	0,0117	0,0006	0,0110 < μ(5) < 0,0124
6	0,0061	0,0004	0,0056 < μ(6) < 0,0066
7	0,0027	0,0002	0,0025 < μ(7) < 0,0029

Wybrane przeze mnie próby były złożone z częściowo różnych leksemów, ale o zbliżonej ich liczbie. Jak widzimy w tabeli 9, przedziały ufności są niewielkie w porównaniu ze średnimi częstościami względnymi. Tak więc niezależnie od próby z pewną dokładnością otrzymamy zbliżone rozkłady względnych częstości. Nasuwa się wniosek, że doboru próby można dokonać w sposób dowolny.

Zbieżność rozkładów częstości

Kolejne pytanie, na które odpowiem w tym artykule, dotyczy zbieżności rozkładu częstości względnych do pewnego rozkładu prawdopodobieństw, czyli

$$\lim_{n \to W} R(x) = P(x), \ (7)$$

gdzie *n* oznacza liczbę badanych leksemów, *W* liczbę leksemów słownika, *R(x)* to względna częstość leksemu o *x* znaczeniach, *P(x)* prawdopodobieństwo wystąpienia w słowniku leksemu o *x* znaczeniach.

Jeśli odpowiedź na powyższe pytanie będzie twierdząca, wnioskiem będzie, że nie jest konieczne badanie całego słownika, a jedynie dostatecznie dużej próby. Przyjrzyjmy się średnim częstościom prób liczącej około 16 411 leksemów, około

Przedziały ufności służą do szacowania wiarogodności wartości danego parametru, którym w moim badaniu jest średnia.

17 268 leksemów, około 18 410 leksemów, 19 526 leksemów, 19 941 leksemów, 20 104 leksemów oraz całej badanej próby liczącej 20 413 leksemów. Dane umieszczone są w tabeli 10.

Tabela 10. Średnie częstości względne

x	Liczba leksemów									
	16411		17268		18410		19526	19941	20104	20413
	$\overline{R(x)}$	*s*	$\overline{R(x)}$	*s*	$\overline{R(x)}$	*s*	R(x)	R(x)	R(x)	R(x)
1	0,6830	0,0180	0,6871	0,0050	0,6907	0,0051	0,6919	0,6919	0,6903	0,6902
2	0,2001	0,0084	0,1973	0,0033	0,1969	0,0040	0,1961	0,1958	0,1963	0,1961
3	0,0660	0,0050	0,0655	0,0010	0,0646	0,0010	0,0639	0,0639	0,0642	0,0644
4	0,0256	0,0022	0,0257	0,0006	0,0254	0,0006	0,0249	0,0250	0,0254	0,0253
5	0,0124	0,0011	0,0117	0,0006	0,0119	0,0005	0,0115	0,0115	0,0115	0,0116
6	0,0060	0,0006	0,0061	0,0004	0,0059	0,0002	0,0055	0,0057	0,0057	0,0059
7	0,0026	0,0003	0,0027	0,0002	0,0025	0,0001	0,0024	0,0024	0,0024	0,0024

Jeśli porównamy liczby w kolumnach 2, 4, 6, 8, 9, 10 i 11 kolejnych wierszy tabeli 10, to widzimy, że wartości *R(x)* dla każdego *x* ustalają się. Oznacza to, że wystarczy próba licząca około 17 000 leksemów, by znaleźć dobre przybliżenie rozkładu prawdopodobieństwa wylosowania ze słownika leksemu o *x* znaczeniach. Ponadto częstości względne, *R(x)* dla *x* = 1, 2, 4, 5, 6, zawarte w kolumnie 11, znajdują się w odpowiednich przedziałach ufności, podanych w tabeli 9. Częstość względna *R(3)* znajduje się na granicy przedziału ufności, a *R(7)* tuż przy lewym końcu przedziału.

Możemy również zauważyć, przyglądając się wartościom częstości względnych umieszczonym w tabeli 10, że liczba leksemów mniejsza niż 16 000 może nie być wystarczająca, by uzyskać zadowalające wyniki. Weźmy próbkę około 2500 leksemów zaczynających się na literę *a*.

Tabela 11. Częstości leksemów *x*-znaczeniowych zaczynających się na literę *a*

x	Częstość	Częstość względna, R(x)
1	1985	0,7579
2	429	0,1638
3	123	0,0470
4	57	0,0218
5	11	0,0042
Suma	wszystkich leksemów: 2619	≈ 0,995

Jak widzimy w tabeli 11, wartości częstości względnych znacznie odbiegają od wyników umieszonych w tabeli 10. Wielkość próby jest więc niewystarczająca. Liczby umieszczone w tabeli 11 potwierdzają też intuicję, że zbiór leksemów zaczynających się na literę *a* i zawierający leksemy nierodzime w polszczyźnie jest słabiej nasycony jednostkami wieloznacznymi.

By dodatkowo potwierdzić powyższe spostrzeżenia, do analizy dodałam jeszcze leksemy zaczynające się na literę *w*. W tabeli 12 umieściłam częstości i częstości względne dla tej nowej powiększonej próby oraz częstości względne dla próby liczącej 17 268 leksemów.

Tabela 12. Częstości leksemów *x*-znaczeniowych

x	Częstość	Częstość względna, R(x)	Średnia częstości względnych dla próby około 17 268 leksemów
1	16961	0,6811	0,6871
2	4953	0,1989	0,1973
3	1670	0,0671	0,0655
4	674	0,0271	0,0257
5	303	0,0122	0,0117
6	159	0,0064	0,0061
7	67	0,0027	0,0027
Suma	wszystkich leksemów: 24 902	≈ 0,995	≈ 0,996

Jeśli porównamy kolumny 3 i 4, widzimy, że otrzymane liczby są bardzo zbliżone. Ponadto, jeśli *R(x)* umieszone w kolumnie 3 porównamy z przedziałami ufności zawartymi w tabeli 9, zauważymy, że dla $x = 2, 5, 6, 7$ *R(x)* znajdują się w tych przedziałach. Natomiast *R(1)* jest tuż przy lewym końcu odpowiedniego dla tej wartości przedziału, a *R(3)* i *R(4)* leżą tuż przy prawym końcu przedziału. Należy pamiętać, iż 95% przedziały ufności oznaczają 95% pewność, że średnie wartości częstości względnych prób o liczebności około 17 000 leksemów znajdą się w tych przedziałach. Wnioskiem z tego eksperymentu jest następujący: powiększenie liczby badanych leksemów nie wnosi nowych informacji do wyników badań.

Podsumowanie

Do weryfikacji prawa Kryłowa wystarczy losować dowolną próbę liczącą około 17 000 leksemów. Uzyska się wtedy dobre przybliżenie rozkładu prawdopodobieństwa wylosowania ze słownika leksemu o *x* znaczeniach. Mając tak dobraną próbę sprawdzić należy własność drugą[4] funkcji *f*, opisanej powyżej. Można to zrobić na dwa sposoby.

Po pierwsze można analizować 99,5% leksemów o znaczeniach *x* mniejszych od pewnej liczby. Jest to metoda weryfikacji prawa Kryłowa zaadoptowana w tym artykule. W wypadku opisanych badań, liczba znaczeń *x* przyjmuje wartości mniejsze lub równe liczbie 7.

Drugi sposób weryfikacji prawa to zastosowanie współczynnika korelacji rangowej Spearmana (zob. szczegółowo: Kułacka 2009). W tabeli 13 umieściłam rangi

[4] Pozostałe własności tej funkcji są oczywiste.

x-znaczeniowych leksemów, rangi dla ich częstości oraz różnice między tymi rangami.

Tabela 13. Rangi dla leksemów o *x* znaczeniach oraz rangi ich częstości

x	Ranga dla x	Częstość	Ranga dla częstości	Różnice między rangami
1	1	14089	1	0
2	2	4004	2	0
3	3	1315	3	0
4	4	516	4	0
5	5	236	5	0
6	6	120	6	0
7	7	50	7	0
8	8	28	8	0
9	9	17	9	0
10	10	10	10	0
11	11	9	11	0
12	12	4	13,5	1,5
13	13	5	12	1
14	14	2	15,5	1,5
15	15	4	13,5	1,5
17	16	1	17,5	1,5
18	17	2	15,5	1,5
27	18	1	17,5	0,5
suma	–	20 413	–	–

Współczynnik korelacji rangowej Spearmana dla powyższych danych jest równy r = 0,9871. Wartość krytyczna dla 18 danych na pięcioprocentowym poziomie istotności dla testu jednostronnego to 0,4000. Wartość r jest wyższa od wartości krytycznej, czyli istnieje dodatnia nieliniowa korelacja między liczbą znaczeń o ich częstością. Najwięcej w słowniku jest leksemów mających jedno znaczenie, a najmniej leksemów wieloznaczeniowych.

Literatura

Altmann G., 2005, *Diversification Processes*, [w:] R. Köhler, G. Altmann, R. G. Piotrowski (red.), *Quantitative Linguistics. An International Handbook*, Berlin, s. 97–113.

Hammerl R., Sambor J., 1993, *O statystycznych prawach językowych*, Warszawa.

Krylov Ju. K., 1982. *Eine Untersuchung Statistischer Gesetzmässigkeiten auf der paradigmatischen Ebene der lexik natürlicher Sprachen*, [w:] BRAK REDAKTORA, *Studies on Zipf's law*, Bochum, s. 234–262.

Kułacka A., 2009, *Warunki zachodzenia prawa Menzeratha-Altmanna*, „LingVaria” nr 1(7), s. 17–28.

SAMBOR J., 1972, *Słowa i liczby. Zagadnienia językoznawstwa statystycznego*, Wrocław.

SAMBOR J., 1989, *O nowym projekcie badań statystycznej struktury słownictwa*, [w:] W. Lubas (red.), *Wokół słownika współczesnego języka polskiego* II, Wrocław, s. 97–113.

SJP: *Słownik języka polskiego PWN*, http://sjp.pwn.pl/ [dostęp: 19 IV 2009].

WIMMER G., ALTMANN G., 2005, *Unified Derivation of Some Linguistic Laws*, [w:] R. Köhler, G. Altmann, R. G. Piotrowski (red.), *Quantitative Linguistics. An International Handbook*, Berlin, s. 791–807.

Procedure of verification of the Krylov Law
Summary

In this article I present the properties of a mathematical model for the frequency distribution function describing the Krylov law. I also outline the uniform procedure to verify the law. I found the sufficient number of lexemes to be gathered for the empirical data to be convergent to a certain function and I point out to how they may be gathered.