

PAWEŁ SWOBODA

Instytut Języka Polskiego PAN, Kraków

REGIONALNE ZRÓŻNICOWANIE ILOŚCIOWE IMIENICTWA W POLSCE

Słowa kluczowe: antroponimia, imiona, geografia nazewnictwa, statystyka leksykalna, bogactwo leksykalne.

STRESZCZENIE

Artykuł dotyczy różnic ilościowych w imiennictwie poszczególnych regionów Polski w latach 1951–2010. Celem było sprawdzenie, czy dane ilościowe dla regionów są zróżnicowane, czy układają się w zwarte obszary geograficzne oraz czy zróżnicowanie międzyregionalne ulegało zmianom w badanym okresie. Do oceny tych różnic użyte zostały wybrane wskaźniki bogactwa leksykalnego stosowane w pracach z zakresu statystyki leksykalnej. Impulsem do podjęcia tego zagadnienia były zaobserwowane wcześniej wyraźne różnice w dystrybucji przestrzennej imion ze względu na ich cechy jakościowe. Badanie wykazało, że także w świetle danych ilościowych występuje czasowe i geograficzne zróżnicowanie w nadawaniu imion w Polsce.

1. WSTĘP

Artykuł stanowi próbę porównawczej analizy wybranych wskaźników dotyczących rozkładu częstości imion nadawanych na obszarze szesnastu współczesnych województw Polski w sześciu równych odstępach czasowych, które wyznaczają dekady: 1951–1960, 1961–1970, 1971–1980, 1981–1990, 1991–2000 i 2001–2010. Celem artykułu jest odpowiedź na następujące pytania: 1) czy występują bądź występowały różnice w strukturze ilościowej imion nadanych w poszczególnych regionach Polski?; 2) jeśli tak, to czy można mówić o pewnych zwartych obszarach, na których dany wskaźnik liczbowy realizuje się w podobny sposób, innymi słowy — czy istnieje w tym zakresie jakaś wyraźna regionalna specyfika, czy też może układ danych jest wszędzie identyczny lub ma charakter przypadkowy?; 3) czy ewentualne zróżnicowanie regionalne uległo w ciągu lat jakimś zmianom, czy też utrzymuje się na stałym poziomie. Impulsem do podjęcia takiej tematyki było wcześniejsze zaobserwowanie wyraźnych regionalnych różnic w zakresie cech jakościowych imion nadawanych w Polsce w latach 1995–2010 (por. Swoboda 2013: 56–67). Okazało się bowiem, że częstość użycia danego imienia bądź grupy imion, mających jakąś wspólną cechę (brzmieniową, genetyczną etc.), może być zależna od czynników społeczno-geograficznych. Pewnych inspiracji i rozwiązań metodologicznych dostarczają również powstałe w ostatnich latach

prace onomastyczne, których autorzy skupili się w znacznej mierze na strukturze statystycznej imiennictwa w poszczególnych krajach i okresach (por. Eshel 2000; Tucker 2001, 2002; Huschka, Gerhards, Wagner 2009; Galbi 2002; Lieberson 2000; Lieberson, Lynn 2003; Mateos, Tucker 2008). Kwestia czasowego i regionalnego zróżnicowania ilościowego była podejmowana wcześniej także w pracach polskich autorów z zakresu imiennictwa (m.in. Gajda 1973, Nowik 1998).

2. METODY I MATERIAŁ

Do zbadania różnic pomiędzy zbiorami imion nadawanych w poszczególnych regionach w ciągu sześciu dekad wykorzystano wybrane klasyczne miary bogactwa leksykalnego, takie jak wielkość słownika, wskaźnik różnorodności Yule'a, wskaźnik stereotypowości Mistríka, wskaźniki koncentracji Lorenza i Mistríka, oryginalności Guirauda, które zastosowała m.in I. Kamińska-Szmaj (1990) w pracy poświęconej różnicom leksykalnym między stylami funkcjonalnymi polszczyzny. Choć proponowanych wskaźników bogactwa leksykalnego jest znacznie więcej (ich obszernego przeglądu dokonano w pracach: Tweedie, Baayen 1998; Pawłowski 2003; Tuldava 2005), postanowiono ograniczyć się tylko do kilku wyżej wymienionych statystyk opisowych, które wydają się wystarczające do wstępnej oceny regionalnego zróżnicowania imiennictwa.

Metody, o których mowa, mają zastosowanie przede wszystkim na gruncie statystyki leksykalnej czy stylometrii, jednak mogą być używane także w przypadku badania zbiorów antroponimów (por. Skowronek 2001). Wymaga to jednak traktowania tych zbiorów jako swego rodzaju *tekstów onomastycznych* czy też *quasi-tekstów* (tamże: 27). Implikuje to z kolei postrzeganie danego regionu (wyznaczonego granicami administracyjnymi współczesnych województw) jako swego rodzaju *quasi-stylu* czy też *odmiany systemu antroponimicznego*, którego reprezentacją są zbiory imion. Zastosowanie metod z zakresu statystyki leksykalnej jest możliwe przede wszystkim dlatego, że duże zbiory imion wykazują podobne cechy rozkładu częstości co teksty języka naturalnego, które określane są mianem *Large Number of Rare Events* (LNRE) (Huschka, Wagner 2010: 17, Baayen 2001: 51).

Badania porównawcze, wykorzystujące do opisu różnic między tekstami wskaźniki ilościowe, wymagają operowania tekstami o możliwie równej długości, gdyż właściwie wszystkie charakterystyki liczbowe tekstu są zależne od jego długości (zob. Tweedie, Baayen 1998: 323 i n., Baayen 2001: 24 i n., Pawłowski 2003: 175). Dotyczy to nie tylko klasycznych miar, które czasem uznawano za niezależne (np. *K* Yule'a, por. Sambor 1969: 79, 112), ale także tych zaproponowanych ostatnio jako niezależne przez I.-I. Popescu¹. Na problem ten w przypadku antroponimii zwrócili uwagę niemieccy

¹ Mowa tu o parametrach *a* i *b* będących ilorzem długości tekstu i odpowiednio liczby *h* (punktu, w którym ranga i frekwencja na liście frekwencyjnej są równe) i liczby *k* (klasa częstości i jej frekwencja w spektrum częstości są równe) (Popescu 2009: 17 i n.). Popescu, badając ich realizacje na tekstach różnej długości w różnych językach, uznał, że nie są one zależne od długości tekstów. Na potrzeby niniejszego artykułu wykonałem testy odporności parametrów *a* i *b* na długość tekstu, losując coraz większe próby imion, two-

naukowcy D. Huschka, J. Gerhards i G. Wagner (2009) w pracy na temat różnic w imiennictwie Niemiec Wschodnich i Zachodnich. Autorzy przyjęli hipotezę o bardziej zindywidualizowanym charakterze imiennictwa w RFN. Zakładali oni, że Niemcy Wschodnie musiały być pod tym względem bardziej zuniformizowane (Huschka i in. 2009: 210). Po przeprowadzeniu obliczeń okazało, że przeciętna liczba użytkowników dla jednego imienia była w RFN dużo wyższa niż w NRD (12,8 : 8,4), co nie tylko przeczyło przyjętej hipotezie, ale nawet dowodziło czegoś zupełnie odwrotnego. Natomiast po zredukowaniu większej próby dla RFN do wielkości próby dla NRD okazało się, że średnia liczba użytkowników jednego imienia była niemal identyczna (8.4 — NRD, 8.3 — RFN), a zatem pomimo różnic jakościowych między dawnymi republikami (w NRD więcej było np. imion słowiańskich itd.) pod względem struktury liczbowej praktycznie niewiele się różniły.

Niestety w wielu dotychczasowych pracach antroponimicznych kwestia równości prób nie była brana pod uwagę i wnioski w nich zawarte mogą być co najmniej dyskusyjne. Uwagi na temat prac zagranicznych autorów, m.in. K. Tuckera (2001) czy A. Eshela (2001), znajdują się w innym artykule Huschki i Wagnera (2010). Ograniczę się do przykładu prac z zakresu polskiej antroponimii. K. Nowik, badając zmiany frekwencyjne w zasobie imion w Polsce do 1994 r., porównywała ponad 19-milionowy zbiór imion z lat 1901–94 z niespełna 3-milionowym zbiorem antroponimów z lat 1981–90. Badaczka zauważyła, że 20 pierwszych imion w pierwszym zbiorze pokrywa ponad 51% sumy użyć imion, a w drugim zbiorze z lat 1981–90 jest to już ok. 66%. Na podstawie tej obserwacji wyciągnęła wniosek, że w latach 80. przy wyborze imienia w większym stopniu kierowano się „modą”, niż miało to miejsce w ciągu całego XX w. (Nowik 1998: 62). Wiadomo tymczasem, że wielkość słownika (liczby użytych wyrazów w tekście) lub jej stosunek do sumy użyć słów w tekście (*type-token ratio*) jest jednym ze wskaźników najbardziej podatnych na długość tekstu (Baayen 2008: 244) i porównanie zbioru imion z lat 80. z pracy K. Nowik z równym liczbowo zbiorem z innego okresu musiałoby dać odmienne wyniki (co nie znaczy, że na pewno przeczyłyby końcowemu wnioskowi autorki). Nawiązując do pracy K. Nowik, podobny błąd powtórzył piszący te słowa, kiedy porównywał z jej wynikami dane polskie z lat 1995–2010 (Swoboda 2013). Z kolei S. Gajda w artykule poświęconym imiennictwu na Opolszczyźnie, porównując ze sobą zbiory różnej wielkości (tj. o sumie użyć np. 1016, 1853, 2319), zaznacza co prawda, że liczba osób w danej grupie wpływa na liczbę użytych imion, jednak za wystarczające do rozwiązania tego problemu uważa wyliczenie średniej liczby użytkowników jednego imienia jako bardziej miarodajnego wskaźnika bogactwa systemu imienniczego (Gajda 1973: 269).

Materiał badawczy niniejszego artykułu stanowią informacje o obywatelach polskich pozyskane z Powszechnego Elektronicznego Systemu Ewidencji Ludności (PESEL), uwzględniającego dane takie jak: imię, płeć, rok urodzenia, miejsce rejestracji

rząc dla nich listy frekwencyjne i spektra częstości, a także obliczając parametry *a* i *b*. Okazało się, że również w ich przypadku istnieje zależność od wielkości próby (długości tekstu).

urodzenia (w formie kodu TERYT²). Materiał podzielono, osobno dla kobiet i mężczyzn, na sześć dekad (od 1951 do 2010 r.). Powodem ograniczenia badania tylko do tych lat był oczywisty fakt, że nie można rozpatrywać materiału imienniczego sprzed 1945 r. w kontekście powojennych podziałów administracyjnych. Z drugiej zaś strony zależało mi na uzyskaniu równych, 10-letnich przedziałów, z których ostatni kończy się na 2010 r. (jest to data graniczna dostępnego mi materiału), dlatego też zrezygnowałem z materiału z lat 1946–50. Materiał z każdej z sześciu dekad podzielono (na podstawie kodu TERYT) na 16 grup odpowiadających województwom. Liczba osób w poszczególnych grupach różniła się znacząco (co wynika z oczywistych różnic demograficznych), a więc i otrzymane zbiory antroponimów musiały różnić się długością. Żeby wyrównać je pod względem długości (tzn. zredukować do długości najkrótszego ze zbiorów), dla każdej z takich 16 grup w sześciu okresach wylosowano bez zwracania 500 podprób liczących po 40 tys. osób³. Liczba 40 tys. podyktowana była tym, że najmniejszy zbiór osób liczył nieco ponad 44 tys. (woj. opolskie, osoby płci żeńskiej, dekada 2001–10)⁴ oraz wygodą obliczeń. Dla każdej z 500 podprób⁵ obliczono wybrane wskaźniki bogactwa leksykalnego (omawiam je w poszczególnych częściach)⁶, których średnie wartości zamieszczono w odpowiednich tabelach⁷. Ponadto zamieszczono w nich (3 wiersz od dołu) wartość średniej dla wszystkich województw, co pozwala zorientować się, jaka była ogólna tendencja dla opisywanego liczbowo zjawiska w danej dekadzie.

Żeby określić stopień zróżnicowania wartości danego wskaźnika między województwami (tzn. czy różnice występujące np. w pierwszej dekadzie były tego samego rzędu co w dekadzie ostatniej), posłużymy się miarą współczynnika zmienności (v), który pozwala na ocenę zmienności cechy, abstrahując od jednostek cechy (por. Hammerl, Sambor 1990: 54). Ujmuje on stosunek odchylenia standardowego do średniej arytmetycznej (dla przejrzystości zwykle mnożony przez 100). Im wyższa wartość współczynnika zmienności, tym różnice pomiędzy województwami były większe. Wartości współczynnika zmienności podawane są w ostatnich dwóch wierszach tabeli.

² Kod TERYT uwzględnia współczesny podział administracyjny, a więc niezależnie od podziału administracyjnego w różnych okresach kod miejsca rejestracji urodzenia został dostosowany do podziału współczesnego.

³ Wzorowałem się na procedurze, którą zastosowali berlińscy badacze przy badaniu różnic w imiennictwie między Niemcami Wschodnimi i Zachodnimi (por. Huschka, Gerhards, Wagner 2009; Huschka, Wagner 2010).

⁴ Dla porównania najwięcej nadań imion — ponad 384 tys. — odnotowano w grupie kobiet w województwie mazowieckim w okresie 1951–1960.

⁵ Okazało się, że stabilne wyniki uzyskiwałem już przy 150 losowaniach, jednak dla zwiększenia pewności postanowiłem pozostać przy 500 losowaniach.

⁶ Podział materiału, losowanie oraz wszystkie obliczenia zostały wykonane przy pomocy pakietu statystycznego R (R Development Core Team 2011).

⁷ Należy pamiętać, że odnoszą się one nie do wszystkich nadań imion w badanym przedziale czasowym, ale do ich 40-tysięcznych podprób.

W drugim z nich pominięto dane dla województwa mazowieckiego, które wykazywało znaczne odchylenia od pozostałych regionów.

W celu przystępniejszej prezentacji uzyskanych wyników dane dla poszczególnych wskaźników umieszczonych w tabelach zostały naniesione także na mapę Polski. Pozwala to na łatwiejszą ocenę tego, czy dane układają się w jakieś zwarte obszary geograficzne, czy też mają układ przypadkowy. Mapy te mają charakter jedynie orientacyjny — skala odcieni jest uzależniona *stricte* od rozpiętości wartości danego wskaźnika w poszczególnych dekadach.

Na koniec należy przypomnieć, że ilościowych wskaźników bogactwa leksykalnego nie można stosować do wartościowania badanych tekstów czy zbiorów imion. Jak zaznacza A. Pawłowski, „pojęcie «bogactwa» jest aksjologicznie nacechowane i konotuje pozytywne atrybuty tekstu (słownictwo ubogie vs bogate). W rzeczywistości, rzeczony wskaźnik wyrażają różne, często trudne do zinterpretowania właściwości tekstu i nie powinny być mechanicznie przekładane na kategorie tradycyjnej stylistyki” (Pawłowski 2003: 176). Dane liczbowe nie orzekają nic o cechach jakościowych jednostek leksykalnych, ich budowie słotwórczej lub semantycznym bogactwie podstaw (Skowronek 2001: 79), a jedynie wskazują na pewne tendencje w zakresie wyboru środków językowych, które możliwe są do zinterpretowania tylko w oparciu o informacje o innych czynnikach, które mogły warunkować strukturę ilościową danej zbiorowości. W niniejszej pracy skupiam się tylko na różnicach ilościowych w zakresie nadawania imion w poszczególnych regionach, nie podejmuję natomiast próby ich wyjaśniania.

3. ANALIZA

3.1. Wielkość słownika

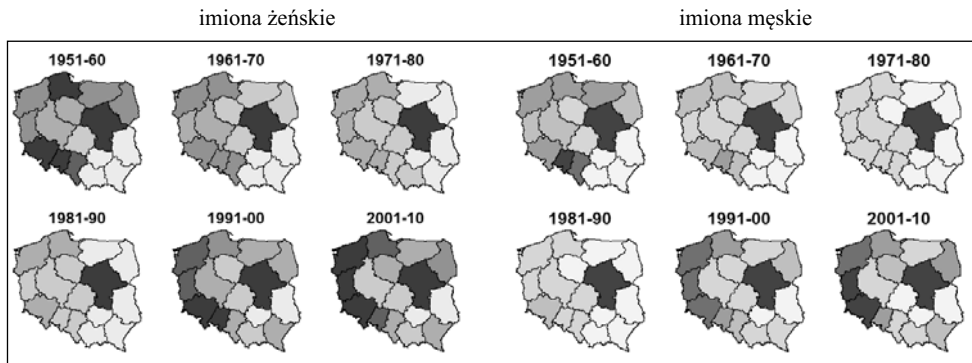
Za najbardziej podstawowy wskaźnik bogactwa leksykalnego uznaje się po prostu wielkość słownika tekstu, a więc liczbę unikatowych typów (wyrazów), które w nim wystąpiły (Tweedie, Baayen 1998: 323 i n.). W naszym przypadku będzie to liczba różnych imion, które wystąpiły w losowych próbach dla poszczególnych województw, w poszczególnych dekadach. W przypadku nierównych prób, jeśli są one zbliżone do siebie rozmiarem, oblicza się średnią częstość wyrazu (imienia) w zbiorze⁸. Biorąc jednak pod uwagę, że dla wszystkich porównywanych grup dysponujemy równymi, 40-tysięcznymi podpróbami, wystarczające jest operowanie liczbą typów. W tabeli 1. zestawione zostały średnie liczby odnotowanych typów w podziale przestrzennym i czasowym.

⁸ Nazywaną też indeksem powtarzalności lub wskaźnikiem iteracji (Sambor 1972, Kamińska-Szmaj 1990: 12).

Tab. 1. Średnia wielkość słownika (liczba typów) imion w województwach, średnia wielkość słownika, współczynnik zmienności (v) między województwami, współczynnik v bez woj. mazowieckiego⁹

woj.	żeńskie						męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	405	400	410	442	632	775	322	342	339	371	559	713
KUJ.	340	345	350	371	480	656	270	279	281	301	446	619
LBL.	295	286	299	312	431	581	242	250	240	262	384	546
LBS.	357	379	395	408	590	754	290	306	301	337	539	745
ŁDZ.	321	333	342	371	482	621	272	289	294	320	428	554
MŁP.	290	312	336	357	484	602	233	260	284	308	446	583
MAZ.	426	476	539	621	669	749	394	467	553	586	610	720
OP.	427	408	383	397	639	728	370	358	325	323	520	649
PDK.	286	302	306	350	514	665	211	237	262	282	454	633
PDL.	372	321	296	332	538	672	296	276	256	293	482	653
POM.	418	401	406	426	589	705	322	328	320	353	496	648
ŚL.	386	389	367	393	531	636	342	337	315	324	472	611
ŚWK.	302	297	300	308	446	565	231	243	256	265	380	563
WRM.	369	348	293	337	522	635	310	287	254	290	446	605
WLP.	343	363	370	402	520	625	283	283	293	332	449	585
ZPM.	373	385	409	422	618	752	303	329	326	367	552	712
śred.	357	359	363	391	543	670	293	304	306	332	479	634
v	13,62	14,12	17,54	18,85	13,30	9,93	17,29	18,71	23,66	22,64	13,29	9,85
v (bez MAZ.)	13,21	11,79	12,63	11,17	12,36	9,84	15,52	13,03	10,77	10,74	11,71	9,56

Rys. 1. Wielkość słownika imion w województwach (dane z tab. 1)*



* Im ciemniejszy odcień, tym większa liczba typów. Każda mapa przedstawia zróżnicowanie w obrębie danej dekady (bez odniesienia do innych okresów).

Z danych zamieszczonych w tabeli 1., naniesionych także na mapę administracyjną Polski, wynika, że najwyższa liczba typów przez pierwszych pięć dekad występowała zawsze w województwie mazowieckim. Fakt ten prawdopodobnie wiąże się z wysoki-

⁹ Rozwinięcie skrótów: dln. — dolnośląskie, kuj. — kujawsko-pomorskie, lbl. — lubelskie, lbs. — lubuskie, łdz. — łódzkie, młp. — małopolskie, maz. — mazowieckie, op. — opolskie, pdk. — podkarpackie, pdl. — podlaskie, pom. — pomorskie, śl. — śląskie, śwk. — świętokrzyskie, wrm. — warmińsko-mazurskie, wlp. — wielkopolskie, zpm. — zachodniopomorskie.

mi wartościami wskaźnika w samej Warszawie, a nie w całym województwie mazowieckim. Poza tym można zauważyć, że poszczególne województwa układają się w zwarte obszary. W pierwszych dwóch dekadach (1951–60, 1961–70) na tle reszty Polski ewidentnie wyróżnia się południowo-wschodnia część Polski (lubelskie, małopolskie, podkarpackie i świętokrzyskie), w której do nominacji osób użyto najmniejszej liczby typów imion (z czasem różnice te uległy niwelacji). Zasadniczo przez cały czas utrzymuje się tendencja, w wyniku której dużo więcej różnych typów imion nadaje się w zachodniej części Polski. Dotyczy to zwłaszcza zewnętrznego pasa województw zachodnich (od opolskiego na południu po pomorskie na północy). W ostatniej, szóstej dekadzie w przypadku imion żeńskich liczbę typów w województwie mazowieckim przewyższają właśnie wskaźniki dla województw: dolnośląskiego, lubuskiego i zachodniopomorskiego, a w przypadku imion męskich wskaźnik dla województwa lubuskiego. Pomimo ciągle utrzymującej się przewagi w liczbie typów imion w województwie mazowieckim oraz w województwach zachodnich, na podstawie analizy współczynnika zmienności (tab. 1), możemy stwierdzić, że różnice te ulegały w ostatnich dekadach stopniowemu zacieraniu, choć jak wynika z zamieszczonych map, ciągle się utrzymują. We wszystkich województwach notuje się (zwłaszcza od połowy badanego okresu) przyrost typów imion, co może świadczyć o ich coraz większej różnorodności. Przyrost ten, poza wskaźnikami dla poszczególnych regionów, obrazują średnie w dekadach zamieszczone w tabeli 1.

3.2. Różnorodność imiennictwa

Do oceny różnorodności słownictwa używany jest często wskaźnik *K* Yule'a, który oblicza się na podstawie list frekwencyjnych wyrazów (imion). Oblicza się go za pomocą wzoru (Sambor 1969: 158):

$$K = \frac{\sum f_i^2 W_i - N}{N^2} \times 10^4,$$

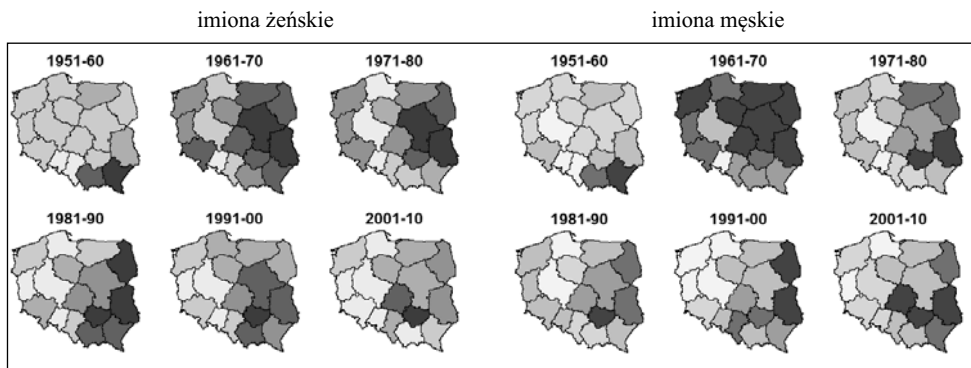
gdzie f_i to częstość o i -tej randze w spektrum częstości, W_i — liczba imion o danej częstości, N — wielkość zbioru (suma użyć wszystkich imion)¹⁰. Im wyższa wartość wskaźnika K , tym niższa różnorodność słownictwa w tekście. Jak pisze I. Kamińska-Szmaj, powołując się na J. Woronczaka, tekst o większym słowniku (liczbie użytych wyrazów) może posiadać mniej różnorodne słownictwo niż tekst o uboższym ilościowo wskaźniku (Kamińska-Szmaj 1990: 14). Porównanie układu danych wartości wskaźnika K z układem liczby typów (wielkości słownika) pozwoli nam na ocenę, czy taki brak zależności ujawnia się także w przypadku materiału imienniczego.

¹⁰ Objasnienie to dotyczy także pozostałych wzorów stosowanych w pracy, w których występuje jeszcze sam symbol W oznaczający liczbę typów (różnych imion), czyli wielkość słownika.

Tab. 2. Wskaźnik różnorodności K Yule'a imion w województwach, średnia wartość K , współczynnik zmienności (v) wskaźnika K między województwami, współczynnik v bez woj. mazowieckiego.

woj.	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	337	314	381	330	234	246	318	302	319	328	261	235
KUJ.	338	301	337	331	241	274	296	321	319	314	269	249
LBL.	353	342	428	364	262	277	325	320	350	346	296	286
LBS.	332	306	371	310	225	250	299	294	319	320	254	228
ŁDZ.	342	320	377	345	257	285	309	315	325	338	279	276
MLP.	402	307	336	350	264	253	385	292	309	319	266	249
MAZ.	334	334	409	343	257	267	300	321	330	335	271	250
OP.	302	244	292	303	223	249	262	234	292	322	267	240
PDK.	440	319	348	350	251	260	411	291	319	329	274	266
PDL.	332	320	391	363	248	278	302	312	330	342	290	270
POM.	331	270	312	312	244	244	297	298	301	303	258	219
ŚL.	295	269	315	321	240	265	285	281	309	319	287	242
ŚWK.	335	323	389	371	274	303	316	295	345	359	287	287
WRM.	357	314	370	317	244	272	313	319	332	321	272	248
WLP.	340	274	309	313	225	247	270	276	290	300	258	240
ZPM.	337	304	374	323	234	262	308	309	319	320	250	234
średnia	344,2	303,8	358,7	334,1	245,2	264,5	312,3	298,8	319,3	325,9	271,2	251,2
v	9,98	8,66	10,85	6,42	6,10	6,22	12,05	7,50	5,22	4,71	4,97	8,02
v bez MAZ.	10,29	8,59	10,65	6,62	6,18	6,44	12,40	7,53	5,34	4,82	5,14	8,30

Rys. 2 Wskaźnik różnorodności K Yule'a w województwach (dane z tab. 2)*



* Im ciemniejszy odcień, tym mniejsza różnorodność imiennictwa. Każda mapa przedstawia zróżnicowanie w obrębie danej dekady (bez odniesienia do innych okresów).

Również w przypadku wskaźnika Yule'a obserwujemy pewne wyraźne regionalne tendencje. Najwyższe wartości odnotowano w pierwszej dekadzie w południowo-wschodniej Polsce (zwłaszcza województwo podkarpackie i małopolskie), a najniższe na Górnym Śląsku (woj. opolskie i śląskie). W późniejszych okresach wyraźny podział ulega zatarciu, choć od trzeciej dekady wyższe wartości wskaźnika K występują raczej w środkowo-wschodniej części Polski (łódzkie, świętokrzyskie, lubelskie, mazowieckie, podlaskie). W ostatnich dwóch dekadach widoczny jest także wyraźny spadek wskaźnika K w stosunku do poprzednich okresów, co oznacza, że mamy w tym okresie do czynienia ze wzrostem różnorodności imiennictwa. Z kolei zróżnicowanie regional-

ne w zakresie wskaźnika różnorodności stopniowo malało, jednak w ostatniej (imiona żeńskie) lub przedostatniej (imiona męskie) dekadzie pogłębiło się.

Geograficzne rozmieszczenie danych dla wskaźnika Yule'a jest zatem nieco odmienne niż w przypadku wielkości słownika, co potwierdza, że liczba unikatowych imion w zbiorze nie musi (i nie może) być traktowana jako główny wskaźnik bogactwa systemu imienniczego.

3.3. Koncentracja imiennictwa

Najczęściej stosowanymi miarami koncentracji w badaniach statystyczno-leksykalnych są wskaźniki Mistríka oraz Lorenza, choć — jak się okaże — nie są one równorzędne.

Wskaźnik Mistríka mierzy stosunek liczby wyrazów o częstościach większych niż 1, a więc wyrazów powtarzających się, do sumy użyć wyrazów (długości tekstu). W niniejszej pracy stosuje się jego zmodyfikowaną postać (Sambor 1972: 223), obejmującą słownictwo o częstościach większych niż 3, który oblicza się według wzoru:

$$I_{konc} = \frac{20W_{i>3}}{N}$$

Wskaźnik I Mistríka może przyjmować dowolne wartości dodatnie. Im wyższa wartość, tym koncentracja niższa. Nie jest to jednak zbyt precyzyjna miara, gdyż mierzy stosunek całości słownictwa mającego częstości większe niż 3 do sumy użyć, bez uwzględnienia zróżnicowania udziału w obrębie poszczególnych klas częstości.

Dużo bardziej wrażliwa i dokładna jest druga ze wspomnianych miar, wskaźnik Lorenza¹¹, który stanowi liczbowy opis nierównomierności podziału ogólnej sumy wartości cechy między poszczególne jednostki zbiorowości (Sobczyk 1998: 56), inaczej nierównomierności rozkładu (Hammerl, Sambor 1990: 82). W przypadku naszych badań mierzenie stopnia koncentracji dotyczy rozkładu częstości jednostek leksykalnych w poszczególnych zbiorach, czyli imion nadawanych w poszczególnych dekadach. Ogólnie rzecz ujmując, większy stopień koncentracji wyrazów w tekście oznacza, że większa część tekstu pokrywana jest przez niewielką liczbę wyrazów najczęstszych. Wskaźnik Lorenza może przybierać wartości z przedziału (0,1). Gdy rozkład jest maksymalnie równomierny (a więc każde imię ma tę samą liczbę użytkowników), wówczas mówimy o braku koncentracji ($L = 0$), natomiast gdy rozkład jest maksymalnie nierównomierny (tj. gdy wszystkie osoby noszą to samo imię), mówimy o zupełnej koncentracji ($L = 1$). Wskaźnik Lorenza oblicza się według wzoru:

$$L = \frac{2 \sum \sum f_i W_i - N}{NW}$$

Zanim przejdziemy do analizy omówionych wskaźników, należy wspomnieć w tym miejscu o pracy S. Liebersona i F.B. Lynn (2003), w której autorzy odnotowują glo-

¹¹ W naukach społecznych i ekonomicznych zwany też współczynnikiem Giniego, a także wskaźnikiem Nierówności Społecznej (często stosowany jest przy opisie nierównomierności rozkładu dochodów).

balną tendencję malejącej koncentracji imiennictwa (badania swoje oparli na danych dla USA, Kanady i licznych krajów europejskich — od Walii po Węgry, jednak bez uwzględnienia Polski). Proces ten miał się zacząć w XIX w. i gwałtownie przyspieszyć w XX w. (Lieberson, Lynn 2003: 241–242). Od razu trzeba wspomnieć, że autorzy nie stosują żadnego z omówionych wyżej wskaźników, lecz za miarę koncentracji uznają pokrycie sumy użyć przez n najczęstszych imion z listy rangowej (zwykle od 10 do 25). Jest to o tyle niefortunna miara, że zwykle będzie zależeć od objętości słownika. Jak zauważyliśmy w poprzedniej części, objętość słownika uległa w ciągu dekad zwiększeniu, w związku z czym procentowy udział częstości w ogólnej sumie nadań najczęstszych arbitralnie określonej liczby imion (np. 10) z dużym prawdopodobieństwem będzie mała (choć oczywiście nie będzie tak zawsze), gdyż maleje także ich udział w słowniku. Tendencja odnotowana przez Liebersona i Lynn dotyczy także polskiego imiennictwa (tab. 3), jednak błędem byłoby traktowanie tej miary na równi z innymi miarami koncentracji.

Tab. 3. Procentowe pokrycie sumy użyć imion w skali ogólnopolskiej przez 10 najczęstszych imion w sześciu dekadach

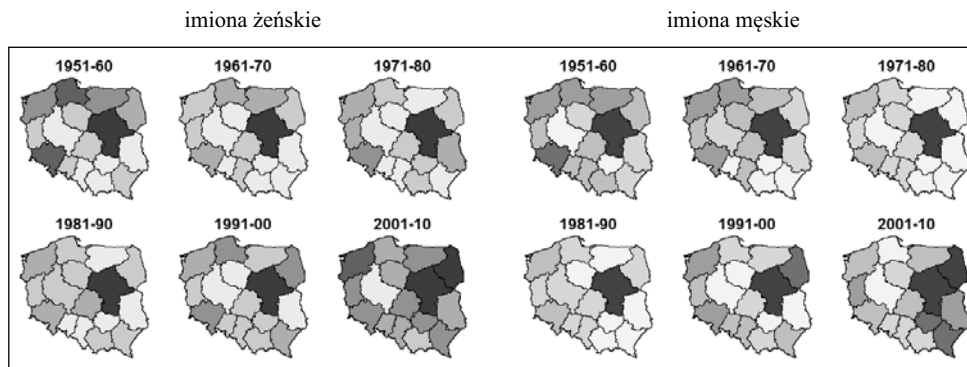
żeńskie						męskie					
1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
50%	46%	50%	49%	40%	42%	46%	45%	47%	48%	41%	39%

Dane dla wskaźnika Lorenza i Mistríka zebrane w poniższych tabelach, poza określeniem stopnia zróżnicowania regionalnego pod kątem koncentracji imion, pozwolą zweryfikować tezę Liebersona i Lynn o malejącej koncentracji.

a) Wskaźnik koncentracji Lorenza

Tab. 4. Wskaźnik koncentracji L Lorenza dla imion w województwach, średnia wartość L , współczynnik zmienności (v) między województwami, współczynnik v bez woj. mazowieckiego.

woj.	żeńskie						męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	0,91	0,90	0,91	0,90	0,91	0,91	0,89	0,89	0,89	0,90	0,91	0,91
KUJ.	0,89	0,89	0,89	0,89	0,89	0,91	0,86	0,88	0,88	0,88	0,90	0,91
LBL.	0,89	0,88	0,90	0,89	0,89	0,91	0,86	0,87	0,87	0,88	0,90	0,91
LBS.	0,89	0,89	0,90	0,89	0,90	0,91	0,87	0,88	0,88	0,89	0,91	0,91
ŁDZ.	0,89	0,89	0,90	0,90	0,90	0,91	0,87	0,88	0,88	0,89	0,90	0,91
MŁP.	0,89	0,88	0,89	0,89	0,90	0,91	0,87	0,86	0,87	0,88	0,90	0,91
MAZ.	0,91	0,92	0,93	0,93	0,92	0,92	0,90	0,92	0,93	0,93	0,92	0,92
OP.	0,89	0,88	0,88	0,88	0,90	0,91	0,87	0,86	0,88	0,89	0,91	0,90
PDK.	0,89	0,89	0,89	0,89	0,90	0,91	0,87	0,85	0,87	0,88	0,91	0,91
PDL.	0,90	0,89	0,89	0,90	0,91	0,92	0,87	0,87	0,87	0,89	0,92	0,92
POM.	0,91	0,89	0,89	0,90	0,91	0,91	0,88	0,89	0,88	0,89	0,91	0,90
ŚL.	0,89	0,89	0,89	0,89	0,90	0,91	0,87	0,88	0,88	0,89	0,91	0,91
ŚWK.	0,89	0,88	0,89	0,89	0,90	0,91	0,85	0,86	0,87	0,88	0,90	0,91
WRM.	0,90	0,89	0,88	0,88	0,90	0,91	0,88	0,88	0,87	0,88	0,91	0,91
WLP.	0,89	0,88	0,88	0,89	0,89	0,90	0,86	0,87	0,87	0,89	0,90	0,90
ZPM.	0,90	0,89	0,90	0,90	0,90	0,92	0,88	0,89	0,89	0,90	0,91	0,91
średnia	0,90	0,89	0,89	0,89	0,90	0,91	0,87	0,88	0,88	0,89	0,91	0,91
v	0,88	1,09	1,28	1,14	0,72	0,39	1,46	1,89	1,68	1,41	0,70	0,52
v bez MAZ.	0,75	0,61	0,79	0,64	0,52	0,35	1,21	1,40	0,89	0,83	0,57	0,48

Rys. 3. Wskaźnik koncentracji Lorentza w województwach (dane z tab. 4)*

* Im ciemniejszy odcień, tym wyższy stopień koncentracji (wg wskaźnika Lorentza). Każda mapa przedstawia zróżnicowanie w obrębie danej dekady (bez odniesienia do innych okresów).

W kontekście wskaźnika Lorentza można mówić o dosyć silnej koncentracji imiennictwa w całym badanym okresie, we wszystkich regionach — jego wartość oscyluje wokół 0,90, choć oczywiście w przypadku niektórych województw występują odstępstwa (wyższe wartości w woj. mazowieckim, niższe w woj. wielkopolskim). Zasadniczo jednak układ danych nie pozwala na zaobserwowanie jakichś zwartych obszarów, na których stale wartości L są niższe lub wyższe. O nikłym zróżnicowaniu świadczą także bardzo niskie wartości współczynnika zmienności, który w przeciągu całego okresu wykazywał do tego tendencję spadkową. Do połowy okresu występowała różnica w koncentracji imion męskich (niższa) i żeńskich, która z czasem również uległa wyrównaniu, co możemy zaobserwować śledząc zarówno wskazania dla poszczególnych województw, jak i wartości ich średniej. Obserwacja zmian w wartościach w sześciu dekadach odzwierciedla jeszcze jedną istotną kwestię — wbrew przywołanym wcześniej twierdzeniom Lieberzona i Lynn, w świetle wskaźnika Lorentza, koncentracja imiennictwa nie tylko nie zmniejsza się, ale nawet wykazuje niewielki wzrost.

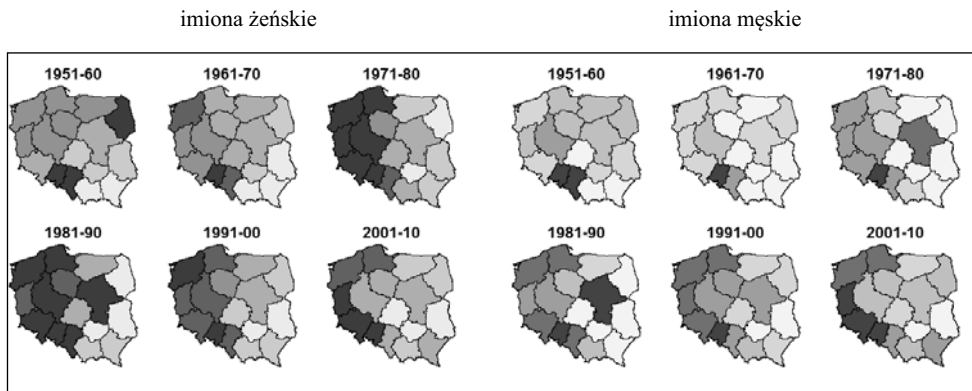
b) Wskaźnik koncentracji Mistríka (z uwzględnieniem modyfikacji J. Sambor)

Tab. 5. Wskaźnik koncentracji Mistríka w województwach, średnia wartość wskaźnika I_{konc} , współczynnik zmienności (v) między województwami, współczynnik v bez woj. mazowieckiego.

woj.	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	0,081	0,088	0,090	0,092	0,114	0,128	0,068	0,068	0,073	0,078	0,093	0,118
KUJ.	0,083	0,082	0,083	0,087	0,107	0,114	0,071	0,064	0,065	0,069	0,087	0,105
LBL.	0,076	0,073	0,074	0,074	0,086	0,105	0,067	0,062	0,062	0,062	0,072	0,095
LBS.	0,083	0,088	0,089	0,088	0,114	0,129	0,071	0,069	0,073	0,074	0,092	0,116
ŁDZ.	0,078	0,083	0,078	0,080	0,097	0,104	0,064	0,065	0,064	0,064	0,078	0,097
MLP.	0,072	0,075	0,078	0,078	0,096	0,109	0,063	0,063	0,066	0,069	0,082	0,105
MAZ.	0,081	0,083	0,080	0,090	0,098	0,113	0,070	0,068	0,074	0,081	0,087	0,104
OP.	0,095	0,099	0,090	0,092	0,123	0,129	0,091	0,090	0,079	0,080	0,102	0,119

woj.	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
PDK.	0,069	0,071	0,076	0,078	0,097	0,115	0,059	0,061	0,063	0,065	0,079	0,108
PDL.	0,091	0,076	0,069	0,073	0,094	0,108	0,071	0,066	0,064	0,063	0,077	0,105
POM.	0,084	0,085	0,090	0,092	0,111	0,120	0,071	0,069	0,070	0,075	0,093	0,112
ŚL.	0,091	0,092	0,086	0,091	0,111	0,114	0,086	0,080	0,072	0,075	0,088	0,109
ŚWK.	0,076	0,077	0,072	0,071	0,085	0,101	0,066	0,064	0,063	0,062	0,071	0,092
WRM.	0,083	0,083	0,077	0,080	0,101	0,109	0,072	0,064	0,064	0,067	0,081	0,100
WLP.	0,085	0,087	0,089	0,093	0,111	0,114	0,075	0,072	0,071	0,074	0,088	0,104
ZPM.	0,083	0,092	0,093	0,093	0,117	0,121	0,068	0,070	0,071	0,076	0,095	0,114
średnia	0,082	0,083	0,082	0,085	0,104	0,115	0,071	0,068	0,068	0,071	0,085	0,106
v	8,20	9,17	9,16	9,33	10,64	7,61	11,30	10,73	7,45	8,97	10,19	7,65
v bez MAZ.	8,47	9,49	9,44	9,52	10,86	7,86	11,69	11,10	7,39	8,49	10,54	7,90

Rys. 4. Wskaźnik koncentracji Mistríka imion w województwach (dane z tab. 5)*



* Im ciemniejszy odcień, tym wyższy stopień koncentracji (wg wskaźnika Mistríka). Każda mapa przedstawia zróżnicowanie w obrębie danej dekady (bez odniesienia do innych okresów).

W odróżnieniu od wskaźnika Lorenza, w przypadku wskaźnika koncentracji Mistríka możemy mówić o wyraźnym (i innym) rozkładzie geograficznym jego wartości, co świadczy o tym, że nie można traktować tych miar jako równorzędnych i stosować ich zamiennie. Poza pierwszą dekadą wyższe liczbowo wskaźniki osiągają województwa w zachodniej części kraju (widoczne jest to zwłaszcza w przypadku imion żeńskich). Zwracają uwagę, zwłaszcza w początkowych dekadach, wysokie wskazania dla województwa opolskiego i śląskiego. Warto odnotować jest to, że w odróżnieniu od pozostałych miar tutaj stosowanych zróżnicowanie między województwami (także z wyłączeniem mazowieckiego) raczej pogłębiało się w ciągu lat (współczynnik zmienności dopiero w ostatniej dekadzie był niższy w stosunku do okresu poprzedzającego). Ponadto wartości wskaźnika Mistríka wykazują tendencję rosnącą, co oznacza malejącą koncentrację. Widzimy zatem, że w zależności od przyjętej metody mierzenia koncentracji można uzyskać zupełnie inne wyniki. Różnica między wynikami dla wskaźnika Lorenza i Mistríka wynika zapewne z tego, że ten pierwszy bierze pod uwagę zniuanso-

wanie rozkładu poszczególnych elementów na liście frekwencyjnej, natomiast wskaźnik Mistrika traktuje *en bloc* wszystkie wyrazy o frekwencji powyżej 3 jako częste.

3.4. Stereotypowość imiennictwa

Stereotypowość leksyki mierzy się m.in. przeciętną długością odstępów między dwoma kolejnymi użyciami wyrazów powtarzających się, a więc mających częstości większe niż 1 (Sambor 1972: 223). Do wyznaczania tej długości służy najczęściej wskaźnik stereotypowości Mistrika wyrażany wzorem:

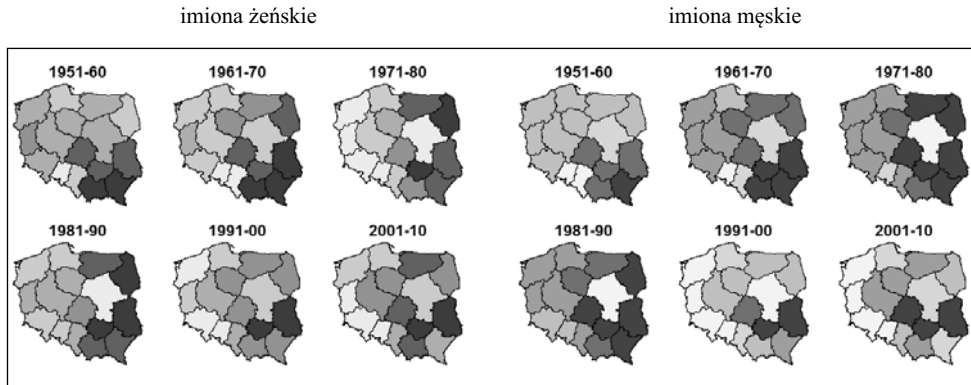
$$I_{stereot} = \frac{N - W_1}{W_{i>1}}$$

Interpretacja tego wskaźnika jest następująca: im wyższa jest jego wartość, tym większy jest stopień stereotypowości imiennictwa i odwrotnie — niższe wartości wskaźnika świadczą o większej oryginalności i zróżnicowaniu ilościowym imiennictwa. W tabeli 6. przedstawione zostały wartości wskaźnika stereotypowości Mistrika, które zostały także naniesione na mapy na rys. 7.

Tab. 6. Wskaźnik stereotypowości w województwach, średnia wartość $I_{stereot}$, współczynnik zmienności (v) między województwami, współczynnik v bez woj. mazowieckiego.

woj.	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	181,6	169,2	162,7	153,9	120,1	102,9	224,2	216,9	208,2	189,8	142,9	112,8
KUJ.	185,6	184,3	185,5	172,2	139,8	118,9	221,4	231,2	232,8	214,5	164,7	129,3
LBL.	205,6	210,2	204,1	204,1	166,2	130,7	245,4	249,5	250,2	243,9	199,9	143,9
LBS.	187,1	171,6	164,9	162,7	123,2	105,6	217,1	219,7	213,3	200,7	147,9	113,4
ŁDZ.	197,9	190,9	190,0	182,5	147,2	125,3	244,9	239,1	236,9	228,6	179,9	141,6
MŁP.	214,0	202,1	194,4	187,2	148,0	125,6	258,3	247,6	229,3	217,0	167,0	131,1
MAZ.	175,6	166,5	156,6	136,7	126,7	112,2	196,2	183,4	161,1	145,1	138,5	119,9
OP.	151,4	151,9	160,4	157,6	113,3	105,3	170,6	167,8	190,8	191,5	139,0	114,4
PDK.	220,5	207,3	203,0	189,1	142,2	116,6	280,6	257,0	241,4	233,0	169,5	122,1
PDL.	167,9	193,4	217,3	201,0	143,7	123,6	217,0	233,9	240,2	234,8	168,5	125,2
POM.	173,4	170,2	167,0	157,2	125,9	110,8	212,8	211,2	216,2	196,8	149,0	122,9
ŚL.	165,2	160,9	173,1	162,8	132,2	118,5	179,8	187,7	207,9	204,4	156,8	126,7
ŚWK.	201,1	199,2	208,0	211,9	163,4	134,6	251,5	251,1	247,6	243,5	198,1	143,6
WRM.	181,7	184,9	200,2	191,1	140,2	124,8	211,5	229,1	249,2	223,9	167,0	132,2
WLP.	183,1	171,2	170,3	161,8	132,8	121,5	214,6	214,6	216,8	202,1	160,2	132,6
ZPM.	181,8	166,7	159,7	156,3	119,5	110,0	223,6	214,3	212,1	194,8	143,5	116,3
średnia	185,8	181,3	182,3	174,3	136,5	117,9	223,1	222,1	222,1	210,3	162,0	126,8
v	9,84	9,76	11,09	12,31	11,15	8,00	12,71	11,57	10,78	12,02	11,71	8,17
v bez MAZ.	10,04	9,80	10,70	11,11	11,32	8,14	12,63	10,84	8,04	8,84	11,33	8,29

Rys. 5. Wskaźnik stereotypowości Mistríka w województwach (dane z tab. 6)*



* Im ciemniejszy odcień, tym imiennictwo jest bardziej stereotypowe. Każda mapa przedstawia zróżnicowanie w obrębie danej dekady (bez odniesienia do innych okresów).

Z przedstawionych danych wynika, że wyższym stopniem stereotypowości w ciągu całego badanego okresu odznaczały się województwa we wschodniej i centralnej części kraju (wyjąwszy woj. mazowieckie), choć początkowo dotyczyło to tylko południowej części wschodniego pasa. Najniższe wartości wskaźnika stereotypowości były notowane (poza woj. mazowieckim) zwłaszcza w województwie dolnośląskim i opolskim, a w ostatnich dwóch dekadach można mówić już o całym pasie sąsiadujących ze sobą województw — od opolskiego do zachodniopomorskiego. We wszystkich regionach dała się zaobserwować ogólna tendencja spadkowa w zakresie stereotypowości imiennictwa, co ilustrują malejące z dekady na dekadę wartości średniej. Zróżnicowanie między poszczególnymi regionami w zakresie wskaźnika stereotypowości w perspektywie całego sześćdziesięciolecia również uległo zmniejszeniu. Warto zauważyć, że układ danych na załączonych mapach jest bardzo podobny do tego, jaki zaobserwaliśmy w przypadku danych o wielkości słownika, co może świadczyć o zależności tych dwóch wskaźników.

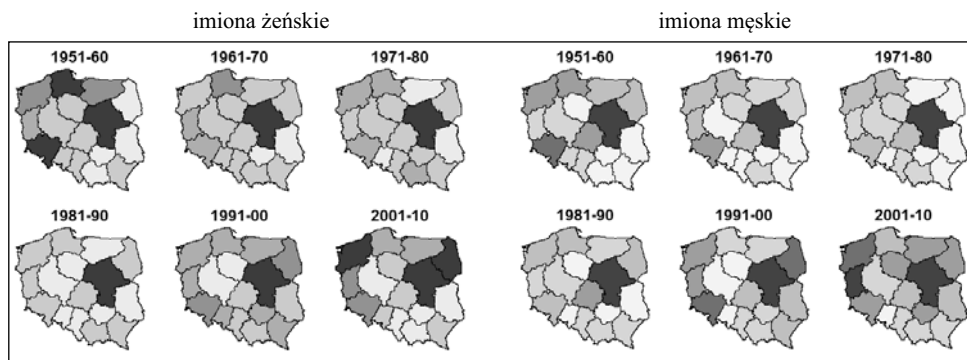
3.5. Oryginalność imiennictwa

Oryginalność słownictwa mierzona jest za pomocą wskaźnika oryginalności Guirauda, obliczanego według wzoru: W_1/W . Mierzy on stosunek słownictwa o częstotliwości $f=1$ do całości słownictwa użytego w tekście. Im wyższa wartość wskaźnika Guirauda, tym słownictwo jest bardziej oryginalne, gdyż większy w nim udział słownictwa rzadkiego, niepowtarzającego się.

Tab. 7. Średnie wartości wskaźnika oryginalności imiennictwa w województwach, średnia ze średnich dla województw, współczynnik zmienności (ν) wielkości między województwami, współczynnik ν bez woj. mazowieckiego.

woj.	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	0,46	0,41	0,40	0,41	0,48	0,50	0,45	0,46	0,43	0,43	0,50	0,51
KUJ.	0,37	0,37	0,38	0,38	0,41	0,49	0,33	0,38	0,39	0,38	0,46	0,50
LBL.	0,34	0,34	0,35	0,37	0,44	0,48	0,33	0,36	0,33	0,37	0,48	0,49
LBS.	0,40	0,39	0,39	0,40	0,45	0,50	0,37	0,41	0,38	0,41	0,50	0,53
ŁDZ.	0,37	0,37	0,39	0,41	0,44	0,49	0,40	0,42	0,43	0,45	0,48	0,49
MŁP.	0,36	0,37	0,39	0,40	0,44	0,47	0,34	0,38	0,39	0,40	0,47	0,48
MAZ.	0,47	0,50	0,53	0,53	0,53	0,53	0,48	0,54	0,55	0,53	0,53	0,54
OP.	0,38	0,36	0,35	0,36	0,45	0,48	0,37	0,34	0,36	0,35	0,45	0,46
PDK.	0,37	0,36	0,36	0,40	0,46	0,49	0,33	0,34	0,37	0,39	0,48	0,49
PDL.	0,36	0,36	0,38	0,40	0,49	0,52	0,38	0,38	0,35	0,42	0,51	0,51
POM.	0,45	0,42	0,41	0,40	0,46	0,49	0,42	0,42	0,42	0,43	0,46	0,50
ŚL.	0,37	0,36	0,37	0,38	0,43	0,47	0,35	0,37	0,39	0,40	0,46	0,49
ŚWK.	0,34	0,33	0,36	0,39	0,45	0,48	0,31	0,34	0,37	0,38	0,47	0,51
WRM.	0,40	0,38	0,32	0,38	0,46	0,50	0,39	0,39	0,37	0,39	0,47	0,50
WLP.	0,36	0,36	0,37	0,39	0,42	0,48	0,34	0,34	0,37	0,40	0,45	0,49
ZPM.	0,41	0,38	0,39	0,40	0,46	0,52	0,41	0,43	0,42	0,44	0,50	0,52
średnia	0,39	0,38	0,38	0,40	0,45	0,49	0,37	0,39	0,40	0,41	0,48	0,50
ν	10,34	10,48	11,83	9,56	6,18	3,64	13,00	13,33	13,04	10,12	4,94	3,87
ν bez MAZ.	9,27	6,49	6,50	3,92	4,42	3,25	10,94	9,92	7,90	6,80	4,23	3,40

Rys. 6. Różnice wskaźnika oryginalności Guirauda imion w województwach (dane z tab. 7)*



* Im ciemniejszy odcień, tym imiennictwo jest bardziej oryginalne. Każda mapa przedstawia zróżnicowanie w obrębie danej dekady (bez odniesienia do innych okresów).

W przypadku wskaźnika oryginalności Guirauda nie można mówić o jakimś charakterystycznym rozmieszczeniu geograficznym jego wartości, trudno jednak nie zauważyć, że dane te mają bardzo podobny układ do tych obliczonych dla wskaźnika koncentracji Lorenza (niższe wskazania dla woj. wielkopolskiego i kujawsko-pomorskiego oraz rosnące wartości w woj. podlaskim), co może świadczyć o zależności między tymi dwoma wskaźnikami. Analiza wartości wskaźnika Guirauda dla poszczególnych województw wskazuje na stopniowy wzrost oryginalności imiennictwa w ciągu badanego okresu. Z kolei malejący współczynnik zmienności świadczy o zacieraniu się różnic między regionami w tym zakresie.

4. PODSUMOWANIE

Przedstawione w artykule dane dotyczące wybranych wskaźników bogactwa leksykalnego obliczonych dla zbiorów imion częściowo dowodzą, że w Polsce występowało bądź występuje geograficzne zróżnicowanie w strukturze ilościowej imiennictwa. W przypadku większości analizowanych statystyk (poza wskaźnikiem koncentracji Lorenza i różnorodności Guiruada) można mówić o wyraźnych zwartych obszarach obejmujących południowo-wschodnią Polskę (woj. małopolskie i podkarpackie czy świętokrzyskie i lubelskie) czy też pas województw obejmujących tzw. Ziemię Zachodnie (opolskie, dolnośląskie, lubuskie, zachodniopomorskie), przyłączone do Polski po 1945 roku. Zasadniczo daje się zaobserwować podział na wschodnią i zachodnią część kraju. Z pewnością za charakterystyczne należy uznać również bardzo częste odstępstwa woj. mazowieckiego od pozostałych regionów, na co, jak już wspomniano, może rzutować specyfika stołecznego miasta. W dużym stopniu zarysowane obszary pokrywają się z tymi, które ujawniły się w trakcie analizy jakościowej materiału imienniczego z okresu 1995–2010 (Swoboda 2013).

W przeciągu całego badanego okresu wytworzyła się tendencja do niwelacji regionalnych różnic — w przypadku wszystkich wskaźników (z wyjątkiem K Yule'a) współczynnik zmienności obrazujący stopień zróżnicowania w zakresie danego wskaźnika między regionami w ostatniej dekadzie był zawsze mniejszy w stosunku do okresów poprzedzających.

Jak wspomniałem na wstępie, w artykule odstąpiłem od próby wyjaśniania opisanych wyżej zjawisk — jego celem było jedynie sprawdzenie, czy zjawiska takie w ogóle występują. Nieprzypadkowy układ geograficzny poszczególnych danych dotyczących liczbowego opisu imiennictwa może być, jak sędzę, warunkowany licznymi czynnikami kulturowymi, społecznymi, demograficznymi oraz językowymi, co powinno stać się kolejnym etapem badania zróżnicowania imiennictwa w Polsce.

LITERATURA

- Baayen R. H. 2001: *Word frequency distributions*, Dordrecht–Boston: Kluwer Academic.
- Eshe l A. 2001: On the Frequency Distribution of First Names, *Names* 49, 55–60.
- Gajda S. 1973: Socjologia imion, *Studia Śląskie* XXIII, 267–296.
- Galbi D.A. 2002: *Long-term trends in personal given name frequencies in the UK*. <<http://www.galbi-think.org/names.htm>> [25 lutego 2015].
- Hammer l R., Sambor J. 1990: *Statystyka dla językoznawców*, Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Huschka D., Gerhards J., Wagner G.G. 2009: Naming Differences in Divided Germany, *Names* 57, 208–228.
- Huschka D., Wagner G.G. 2010: *Statistical Problems and Solutions in Onomastic Research — Exemplified by a Comparison of Given Name Distributions in Germany Throughout the 20th Century*, SOEPpaper No. 332, Berlin.
- Kamińska-Szmaj I. 1990: *Różnice leksykalne między stylami funkcjonalnymi polszczyzny pisanej. Analiza statystyczna na materiale słownika frekwencyjnego*, Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

- Liebertson S. 2000: *A Matter of Taste. How Names, Fashions, and Culture Change*, New Haven: Yale University Press.
- Liebertson S., Lynn F.B. 2003: Popularity as a taste: an application to the naming process, *Onoma* 38, 235–276.
- Mateos P., Tucker K. 2008: Forenames and Surnames in Spain in 2004, *Names* 56, 165–184.
- Nowik K. 1998: Zmiany frekwencyjne w zasobie imion w Polsce powojennej, [w:] E. Jakus-Borkowa, K. Nowik (red.), *Najnowsze przemiany nazewnicze*, Warszawa: Energeia, 57–71.
- Pawłowski A. 2003: O problemie atrybucji tekstu w lingwistyce kwantytatywnej (na przykładzie tekstów Romaina Gary), [w:] J. Linde-Usiekniewicz, R. Huszcza (red.), *Prace językoznawcze dedykowane Profesor Jadwidze Sambor*, Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 169–190.
- Popescu I.-I. in. 2009: *Word Frequency Studies*, Berlin–New York: Mouton de Gruyter.
- R Development Core Team 2011: R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing. Available online at <http://www.R-project.org/>.
- Sambor J. 1969: *Badania statystyczne nad słownictwem (na materiale „Pana Tadeusza”)*, Wrocław: Zakład Narodowy im. Ossolińskich.
- Sambor J. 1972: *Słowa i liczby. Zagadnienia językoznawstwa statystycznego*, Wrocław: Zakład Narodowy im. Ossolińskich.
- Skowronek K. 2001: *Współczesne nazwisko polskie. Studium statystyczno-kognitywne*, Kraków: Wydawnictwo Naukowe DWN.
- Sobczyk M. 1998: *Statystyka. Podstawy teoretyczne. Przykłady — zadania*, Lublin: Wydawnictwo Uniwersytetu M. Curie-Skłodowskiej.
- Swoboda P. 2013: Imiona częste w Polsce w latach 1995–2010 oraz ich zróżnicowanie w czasie i przestrzeni, *Onomastica* LVII, 19–69.
- Tucker K.D. 2001: Distribution of forenames, surnames, and forename-surname pairs in the United States, *Names* 49, 69–96.
- Tucker K.D. 2002: Distribution of Fornames, Surnames, and Forename-Surname Pairs in Canada. *Names* 50, 105–132.
- Tuldava J. 2005: Stylistics, author identification, [w:] R. Köhler, G. Altmann, R.G. Piotrowski (red.), *Quantitative Linguistik: Ein internationales Handbuch / Quantitative Linguistics: An International Handbook*, Berlin: Walter de Gruyter, 368–387.
- Tweedie F., Baayen R.H. 1998: How variable may a constant be? Measures of Lexical Richness in Perspective, *Computers and the Humanities* 32, 323–352.

SUMMARY

Regional quantitative differentiation of first names in Poland

Key words: anthroponymy, first names, onomastic geography, lexical statistics, lexical richness.

The article examines the quantitative differentiation in naming in the various Polish regions during the period of 1951–1960. The aim was to check whether the quantitative data for names are regionally differentiated, and has the interregional differences changed during the period considered. Selected measures of lexical richness (originally applied in the field of lexical statistics) were used to evaluate these differences. The direct reason for considering this case were the previously observed clear spatial patterns of first names with regard to its qualitative features. Present study has shown that also in the case of quantitative data there are some temporal and spatial differentiation in naming practices in Poland.