

INSTYTUT JĘZYKA POLSKIEGO POLSKIEJ AKADEMII NAUK

Paweł Swoboda

**Imiona w Polsce w XX w. i na początku XXI w.
w ujęciu statystycznym i socjolingwistycznym**

Rozprawa doktorska

Promotor:

prof. dr hab. Katarzyna Skowronek

Kraków 2019

Spis treści

I. Wstęp	1
I.1. Cel, założenia i zakres pracy	1
I.2. Struktura pracy	5
I.3. Materiał badawczy.....	7
I.3.1. Imiona w bazie PESEL – charakterystyka materiału badawczego	7
I.3.2. Imiona w bazie PESEL na tle innych rejestrów w dotychczasowych badaniach imiennictwa.....	15
II. Zagadnienia teoretyczno-metodologiczne	21
II.1. Stan badań	21
II.1.1. Zarys stanu badań nad polskim imiennictwem.....	21
II.1.2. Socjolingwistyczne aspekty badań nad współczesnym imiennictwem	27
II.1.3. Metody statystyczne w badaniach nad imiennictwem.....	44
II.2. Założenia metodologiczne.....	64
III. Parametry rozkładu częstości wyrazów w zbiorach imienniczych	72
III.1. Zagadnienie wielkości i równości korpusów imienniczych	75
III.2. Skalowanie zbiorów imion, wielkość i liczba próbek	81
III.3. Wyniki analizy parametrów rozkładu częstości.....	84
III.3.1. Wielkość słownika	84
III.3.2. Liczba hapaks-, dis- i trislegomenów	90
III.3.3. Wskaźnik oryginalności Guirauda	96
III.3.4. Wskaźnik różnorodności K Yule'a.....	102
III.3.5. Wskaźnik stereotypowości Mistríka	107
III.3.6. Parametr α	112
III.3.7. Wskaźnik koncentracji Lorenza	119
III.3.8. Wskaźnik koncentracji Mistríka.....	124
III.3.9. Pokrycie sumy użyć przez r najczęstszych imion	129
III.4. Podsumowanie wyników analizy.....	137
IV. Tempo rotacji	144
IV.1. Tempo rotacji w imiennictwie ogólnopolskim	146
IV.2. Tempo rotacji w imiennictwie na obszarze współczesnych województw.....	154
IV.3. Tempo rotacji w imiennictwie obszarów o różnym stopniu zurbanizowania.....	159
IV.4. Szczegółowe badanie tempa rotacji w skali rocznej – casus II wojny światowej	162
V. Badanie regionalnego zróżnicowania frekwencji względnej imion metodą hierarchicznej analizy skupień	173
V.1. Wprowadzenie	174
V.2. Wybór metryki odległości.....	177
V.3. Dobór cech diagnostycznych	180

V.4. Wybór metody łączenia	183
V.5. Wyniki hierarchicznej analizy skupień – imiennictwo na obszarze współczesnych województw	187
V.6. Wyniki hierarchicznej analizy skupień – imiennictwo na obszarze współczesnych powiatów woj. śląskiego.....	226
VI. Zakończenie.....	232
VII. Bibliografia.....	238

I. Wstęp

I.1. Cel, założenia i zakres pracy

Imię, rozumiane jako indywidualne imię własne, nadawane dziecku z woli rodziców w akcie nominacji (Malec 2001: 8) jest podstawowym, trwałym i powszechnym elementem systemu nazywania jednostek ludzkich, obecnym we wszystkich kulturach i epokach (Barnard, Spencer (red.) 2011: 268–270, Malec 1998: 97). W Polsce, podobnie jak w większości krajów świata, owa kategoria antroponimiczna nie tylko podlega ochronie prawnej, ale jest także prawnie sankcjonowana, co oznacza, że wszyscy obywatele polscy muszą posiadać imię (Tylec 2013: 17–44). Codziennie imiona nadawane są setkom lub nawet tysiącom nowo narodzonych dzieci. Ich rodzice (lub w szczególnych przypadkach urzędnicy) wybierają je spośród zasobu imienniczego istniejącego w danym języku, rzadziej zaś tworzone są (lub adaptowane z innych języków i kultur) imiona nowe (Malec 1998: 98). Od lat 70. XX wieku imię wraz innymi informacjami o poszczególnych obywatelach i mieszkańcach Polski (m.in. takimi jak płeć, rok i miejsce rejestracji urodzenia) zaczęło być zapisywane w Powszechnym Elektronicznym Systemie Ewidencji Ludności (PESEL). Dotyczyło to nie tylko osób urodzonych w tym okresie, ale wszystkich żyjących wówczas osób; system PESEL zawiera zatem informacje o imionach osób urodzonych w latach wcześniejszych, nawet w ostatnim ćwierćwieczu XIX w. Istnienie tak wielkiej bazy danych oraz rozwój narzędzi elektronicznych do ich analizy, stwarza możliwości przeprowadzenia wielkoskalowych badań nad imiennictwem Polaków.

Przedmiotem badania w niniejszej pracy są imiona (pierwsze imiona), których użyto do nominacji osób obywatelstwa polskiego urodzonych w okresie od początku XX w. do 2010 r. Na materiał ten składa się przeszło 50 mln okazów odpowiadających imionom nadanym każdej z osób zarejestrowanej w bazie PESEL, którym towarzyszy informacja o roku i miejscu rejestracji urodzenia oraz płci osoby, której nadano imię. Głównym celem pracy jest zbadanie zróżnicowania (oraz jego zmienności) struktury kwantytatywnej polskiego systemu imienniczego w kilku perspektywach czasowych i przestrzennych. W toku analizy postaram się odpowiedzieć na następujące pytania, które jednocześnie wyznaczają ramy niniejszej pracy: 1) czy w imiennictwie ogólnopolskim na przestrzeni rozpatrywanego okresu występowało czasowe zróżnicowanie w zakresie charakteryzujących go cech ilościowych?; 2) czy przez pryzmat właściwości ilościowych możemy mówić o zróżnicowaniu regionalnym imiennictwa oraz czy zróżnicowanie to odzwierciedla określony typ regionalizacji, a jeśli tak, to czy taka wariancja regionalna ulegała przeobrażeniom w czasie?; 3) czy w zakresie cech

kwantytatywnych występowały bądź występują różnice między imiennictwem ośrodków wielkomijskich i obszarów o niższym poziomie urbanizacji?; 4) czy występują odmienne tendencje ilościowe w nadawaniu imion żeńskich i męskich, a jeśli tak, to czy mają one charakter trwały, czy też ulegały przeobrażeniom w czasie i przestrzeni?

Innymi słowy, chodzi o to, czy zróżnicowanie i zmienność struktury kwantytatywnej lub też charakterystyki liczbowej rozkładu częstości oraz dynamiki zmian w polskim imiennictwie były warunkowane takimi czynnikami jak moment nadania imienia (rok, przedział czasowy), przynależność terytorialna (miejsce urodzenia czy też miejsce zamieszkania rodziców) czy płeć osoby otrzymującej imię, a zatem czynnikami pozajęzykowymi. Przeprowadzone badania pozwolą ustalić, czy w rozpatrywanym okresie można przez pryzmat cech ilościowych wyodrębnić czasowe, regionalne czy społeczne odmiany systemu imienniczego, charakteryzujące się odmiennymi tendencjami w zakresie tych cech.

Przedmiot badania – imiennictwo – sytuuje niniejszą pracę w obrębie onomastyki czy też antroponomastyki, jednak zarysowany powyżej ogólny cel pracy i stawiane w niej pytania implikują odwołanie się w trakcie analizy badanego materiału i interpretacji jej wyników do metod z zakresu lingwistyki statystycznej oraz socjolingwistyki (lub wężej: socjoonomastyki), zasygnalizowanych już w tytule pracy. Samo wykorzystanie narzędzi statystycznych rodzi dodatkowy cel – mianowicie próbę refleksji metodologicznej dotyczącej problematyki ich stosowania, ciągle, jak się wydaje, mało obecnej w polskich badaniach nad imiennictwem.

Wykorzystanie metod z zakresu statystyki leksykalnej do opisu i komparacji zbiorów antroponimicznych wymaga przyjęcia pewnych założeń. Imiona, podobnie jak inne kategorie onimiczne, należą do zasobu leksykalnego języka (Malec 2007: 125), zaś ich zbiory, których granice i zasięg mogą być różnie definiowane, mogą być traktowane jako swego rodzaju abstrakcyjne „teksty onomastyczne” (Skowronek 2001: 26–27). Pojmowane w taki sposób zbiory antroponimiczne, podobnie jak teksty języka naturalnego, stanowią pewne struktury ilościowe, w których każdemu elementowi (np. wyrazowi) odpowiada określona częstość użycia. U podstaw językoznawstwa statystycznego, zwłaszcza statystyki czy stylometrii, leży przekonanie, że frekwencja poszczególnych elementów w języku i jej rozkład może być cechą dystynktywną np. idiolektu, języka grupy społecznej, określonego gatunku tekstu bądź stylu danego autora (Sambor 1972: 13, 19–20, Dunaj 1988: 83, Eder 2014: 92–95). Zbiory imion nadawanych przez rodziców np. w różnych regionach czy przedziałach czasowych, podobnie

jak zbiory wyrazów użytych w tekstach różnych autorów, mogą różnić się między sobą nie tylko cechami jakościowymi, ale także charakterystykami wyrażanymi liczbowo, nie odnoszącymi się wprost do formy językowej imion, a związanymi m.in. z rozkładem częstości imion w zbiorach.

Statystyka oferuje nam narzędzia niezbędne do dokonywania obliczeń i ich analizy w obrębie tak wielkich zbiorów danych jak wspomniany wielomilionowy materiał z bazy PESEL, jednak ze względu na przedstawiony cel i zakres pracy – obejmujący ukazanie stopnia zmienności i zróżnicowania imiennictwa polskiego w aspekcie diachronicznym (czasowym), jak i synchronicznym (przestrzennym, demograficznym) – przy definiowaniu granic porównywanych zbiorów imion konieczne jest odwołanie się do kategorii socjolingwistycznych. Kształt systemu imienniczego, jego rozwój i funkcjonowanie motywowane są bowiem nie tylko czynnikami lingwistycznymi, ale przede wszystkim ekstralingwistycznymi – historycznymi, społecznymi, kulturowymi (Abramowicz 1994: 152). Zbiory imion nadawanych w różnych miejscach i czasie oraz ich wzajemne zróżnicowanie są zatem również kształtowane przez cechy ich twórców, funkcjonujących w określonym kontekście społecznym – zewnętrzne przyczyny wariantywności językowej, jak podaje D. Bartol-Jarosińska za W. Labovem, wywodzą się bowiem z „charakterystyki mówiącego i charakteru używania języka” (Bartol-Jarosińska 1991: 57). Biorąc pod uwagę, że zbiory imion nie mają jednego twórcy, ale są efektem kolektywnych zachowań nazewniczych, za „mówiącego” uznamy w tym wypadku całą społeczność językową czy też wspólnotę komunikatywną, która korzysta z określonego repertuaru językowego, a ściślej – imienniczego. Na podstawie informacji o nosicielach imion w materiale badawczym granice takich hipotetycznych wspólnot możemy wyznaczyć poprzez dwie kategorie: przynależności terytorialnej (osoby dokonujące aktu nominacji np. w danym mieście lub regionie) i/lub przynależności czasowej (osoby dokonujące aktu nominacji w danym roku lub przedziale czasowym), mogące stanowić społeczne osie wariantywności języka (Bartol-Jarosińska 1991: 58–59). O „charakterze używania języka” w obrębie tych wspólnot świadczą z kolei wybrane przez te społeczności środki językowe, w tym przypadku imiona tworzące cały zbiór, które składają się na ich repertuar imienniczy.

Zbiory imion nadanych przez poszczególne społeczności (określane przez czas i/lub miejsce, w którym funkcjonowały), stanowiące swego rodzaju teksty onomastyczne, jak już wcześniej wspomniano, mogą być zróżnicowane pod kątem cech ilościowych, badanych przy użyciu metod statystycznych. Stopień tego zróżnicowania wyznaczony przez realizację danej

cechy związanej z rozkładem częstości elementów w poszczególnych zbiorach lub dynamiką zmian w ich obrębie, posłuży nam do oceny siły lub słabości powiązań pomiędzy poszczególnymi wspólnotami oraz pozwoli na stwierdzenie, czy w zakresie realizacji tej cechy możemy mówić o istnieniu jakiejś czasowej lub geograficznej (regionalnej) odmiany systemu imienniczego, reprezentowanej przez jedną lub więcej wspólnot (tworzących szerszą społeczność).

I.2. Struktura pracy

Praca składa się z siedmiu części: wstępu (I), rozdziału poświęconego zagadnieniom teoretyczno-metodologicznym (II), trzech zasadniczych rozdziałów analityczno-badawczych (III, IV i V), zakończenia (VI) oraz bibliografii (VII).

W części wstępnej przedstawiony został materiał badawczy, który stanowią wybrane dane dotyczące imion i ich nosicieli pozyskane z bazy danych PESEL. Omówiona została jego struktura, a także sposób i zakres jego wykorzystania w pracy (I.3.1). Porównałem go również z innymi rejestrami stosowanymi dotychczas w badaniach imienniczych (I.3.2).

Rozdział II skupia się na kwestiach teoretyczno-metodologicznych. W podrozdziale II.1.1 w zwięzły sposób zarysowano historię i stan badań nad polskim imiennictwem. W sposób bardziej szczegółowy omówione zostały badania antroponimiczne, tak polskie, jak i zagraniczne, o ukierunkowaniu socjolingwistycznym (II.1.2) oraz prace, w których do analizy systemów imienniczych użyto metod statystycznych (II.1.3). W nawiązaniu do wypracowanych w nich metod przedstawiłem założenia metodologiczne niniejszej pracy (II.2). Należy przy tym zaznaczyć, że z uwagi na różny charakter przeprowadzonych analiz poszczególne metody statystyczne wykorzystane w pracy zostały szczegółowo omówione w odpowiednich rozdziałach badawczo-analitycznych.

Rozdział trzeci poświęcony jest porównawczej analizie parametrów rozkładu częstości wyrazów w zbiorach imienniczych w Polsce w XX i XXI w. Zasadnicza część analityczna (III.3) została poprzedzona podrozdziałami dotyczącymi zagadnienia wielkości i doboru korpusów imienniczych, na których zostały przeprowadzone badania (III.1 i III. 2). Na wspomniane parametry rozkładu częstości składają się: wielkość słownika (III.3.1), udział w imiennictwie najrzadszej leksyki (hapaks-, dis- i trislegomenów) (III.3.2), wskaźnik oryginalności Guirauda (III.3.3), wskaźnik różnorodności K Yule'a (III.3.4), wskaźnik stereotypowości Mistríka (III.3.5), parametr α z prawa Zipfa (III.3.6) oraz miary koncentracji – Lorenza (III.3.7), Mistríka (III.3.8) oraz pokrycie sumy użyć wszystkich imion w danym zbiorze przez r najczęstszych imion (III.3.9). Każdemu ze wskaźników poświęcony jest osobny podrozdział, w którym dokonano jego omówienia oraz obliczeń na zbiorach imienniczych (żeńskich i męskich) w kilku perspektywach czasowych i przestrzennych: A) w imiennictwie ogólnopolskim w kolejnych latach z przedziału 1906–2010, B) w imiennictwie na obszarze współczesnych województw w sześciu następujących po sobie dekadach w latach 1951–2010, C) w imiennictwie wybranych ośrodków wielkomiejskich (Wrocławia, Krakowa, Warszawy i Poznania) na tle województw, w których się znajdują

(również w sześciu następujących po sobie dekadach w latach 1951–2010). W podrozdziale III.4 dokonano zbiorczego podsumowania wyników analizy.

W rozdziale czwartym skupiono się na zagadnieniu tempa rotacji w imiennictwie polskim. W pracach poświęconych badaniu dynamiki zmian w imiennictwie tempem rotacji (ang. *turnover rate*) określa się stopień liczby zmian pomiędzy listami frekwencyjnymi *r* najczęstszych imion w wybranych punktach czasowych. Analogicznie do rozdziału poświęconego wskaźnikom bogactwa leksykalnego w rozdziale tym zbadano tempo rotacji (mierzone w odstępach 5- i 10- letnich, pozwalających na uchwycenie długofalowych tendencji) w kilku perspektywach czasowych i przestrzennych: w imiennictwie ogólnopolskim w latach 1900–2010 (IV.1), w imiennictwie na obszarze współczesnych województw w latach 1945–2010 (IV.2), a także w imiennictwie obszarów wielkomiejskich i sąsiadujących z nimi obszarów o niższym stopniu zurbanizowania (IV.3). Dodatkowo przedstawiono wyniki analizy tempa rotacji mierzonego w odstępach rocznych, która ujawniła wyraźne zmiany w imiennictwie związane z okresem II wojny światowej, zasługujące na szczegółowe omówienie (IV.4).

W rozdziale V w celu zbadania regionalnego zróżnicowania w zakresie układu imion na listach frekwencyjnych wykorzystałem metodę hierarchicznej analizy skupień. Jej głównym celem jest ułożenie w grupy obiektów (w tym przypadku zbiorów imion nadawanych na określonym terytorium), reprezentowanych przez określone cechy (w tym przypadku częstość względną poszczególnych imion w tych zbiorach) w taki sposób, aby odzwierciedlić stopień ich powiązania. Badanie zostało przeprowadzone na poziomie województw w kolejnych latach z okresu 1946–2010 oraz powiatów województwa śląskiego w dwudziestoletnich przedziałach z okresu 1931–2010. Części zasadnicze, w których prezentowane są wyniki badania (V.5, V.6), poprzedzone zostały podrozdziałami V.1–4, w których omówione zostały zagadnienia metodologiczne związane z zastosowaniem wspomnianej metody.

Ostatni rozdział (VI) zawiera podsumowanie wyników przeprowadzonych badań. Przedstawione zostały najistotniejsze zjawiska zaobserwowane w toku analizy obejmujące takie zagadnienia, jak charakterystyczne konfiguracje czasowe i geograficzne związane z ilościową strukturą imiennictwa oraz jej zróżnicowanie w obrębie zbiorów imion męskich i żeńskich. W rozdziale tym podjęto również próbę interpretacji zaobserwowanych zjawisk na tle historyczno-społecznym. Zarysowane zostały też dalsze perspektywy badawcze.

I.3. Materiał badawczy

I.3.1. Imiona w bazie PESEL – charakterystyka materiału badawczego

Materiał badawczy niniejszej pracy stanowią dane o osobach obywatelstwa polskiego urodzonych w latach 1901–2010 pozyskane z bazy PESEL, które zostały mi udostępnione w formie plików tekstowych w formacie .csv na potrzeby badawcze przez Ministerstwo Spraw Wewnętrznych.¹ Przykładowy wycinek tych danych prezentuje się następująco:

JULIAN	1	1986	1863011
KAROL	1	1986	2011085
PRZEMYSŁAW	1	1986	1012011
JUSTYNA	2	1986	2861011
MARTA	2	1986	0862011
MALWINA	2	1986	3019011

Każda linia odpowiada jednej osobie. Na rekord składają się następujące informacje zawarte w kolejnych kolumnach: I – imię (pierwsze); II – płeć, zakodowana jako 1 (męska) i 2 (żeńska); III – rok urodzenia; IV – miejsce rejestracji urodzenia w formie siedmiocyfrowego kodu terytorialnego TERC², składającego się z trzech członów: człon I – dwucyfrowy symbol województwa nadany województwom ułożonym w kolejności alfabetycznej; człon II – dwucyfrowy symbol powiatu nadany powiatom danego województwa; człon III – trzycyfrowy symbol gminy i jej rodzaju (miejska, wiejska itd.). Taki zakres danych, ograniczony do informacji o imieniu, płci, roku i miejscu rejestracji urodzenia, był maksimum, które mogłem uzyskać z bazy danych PESEL z uwagi na ochronę danych osobowych. Należy zaznaczyć, że o ile dane dotyczące imienia, płci i roku urodzenia pojawiają się obligatoryjnie, to w przypadku kodu TERC występują luki (szerzej omawiam to w dalszej części tego podrozdziału). Niestety, z uwagi na przepisy o ochronie danych osobowych niemożliwe było pozyskanie bardziej szczegółowych informacji, np. o nazwisku, imionach rodziców i innych.

Jak już wspomniano w części I.1.1 system PESEL zaczął być wdrażany w latach 70. XX wieku, kiedy zaczęto wprowadzać do niego dane o wszystkich obywatelach i mieszkańcach Polski wówczas żyjących, by w kolejnych latach już na bieżąco rejestrować w nim nowo narodzone osoby wraz z ich imionami. Oznacza to, że zawiera on informacje o osobach nie tylko urodzonych po tym okresie, ale także wcześniej, nawet w XIX w. (jeśli wówczas

¹ Obecnie system PESEL prowadzony jest przez Ministerstwo Cyfryzacji.

² TERC – System identyfikatorów i nazw jednostek podziału administracyjnego, będący składnikiem Krajowego Rejestru Urzędowego Podziału Terytorialnego Kraju (TERYT).

jeszcze żyły). Należy jednak mieć świadomość, że odpowiada on populacji generalnej urodzonych (wszystkim osobom, które otrzymały imię w danym roku) dopiero w przypadku danych od roku 1984, natomiast wcześniej możemy mówić jedynie o większych lub mniejszych próbach z populacji. Do kwestii tej powrócę jeszcze, omawiając objętość materiału badawczego, a także w rozdziale III, w częściach poświęconych zagadnieniu wielkości i równości korpusów imienniczych oraz ich skalowaniu (III.1 i III.2).

Materiał badawczy otrzymany z PESEL zawierał pierwotnie informacje o 48 532 010 osobach, którym łącznie nadano 43 883 różne imiona. Przed przystąpieniem do zasadniczych prac konieczne było jednak jego przejrzanie pod kątem ewentualnych błędów. Okazało się, że niektóre informacje dotyczące imion zawierały błędy lub też ich format nie odpowiadał założeniom niniejszej pracy. Po możliwie dogłębnym przeanalizowaniu większości przypadków postanowiłem, że wszystkie rekordy zawierające budzące wątpliwości informacje o imionach zostaną wyeliminowane z materiału badawczego. Zdecydowało o tym kilka czynników. Przede wszystkim liczba takich rekordów wyniosła tylko 14 973, co stanowi zaledwie 0,03% pierwotnej bazy. Pozwoliło to jednak na eliminację 6794 typów (ok. 15,5% pierwotnej liczby), czyli unikatowych imion, których unikatowość w dużej mierze była efektem błędów popełnionych przy wprowadzaniu danych do bazy PESEL. Dotyczyło to zatem formuł imienniczych o bardzo rzadkim, zwykle błędnym/przypadkowym użyciu (średnio każda z takich formuł pojawiała się w całym zbiorze ok. 2,2 raza). Ich korekta byłaby możliwa tylko w pewnej części, natomiast rozstrzygnięcie wątpliwości co do prawidłowej formy imienia w każdym z prawie 15 tys. przypadków o ile nie jest zadaniem niemożliwym, to z pewnością bardzo trudnym. Ostatecznie postanowiono, że z analizy zostaną wyeliminowane wszystkie rekordy, które w polu „imię” zawierają znaki inne niż znaki polskiego alfabetu (łącznie z literami Q, V i X) oraz następujące znaki diakrytyczne z innych alfabetów łacińskich: Á Â Ã Ä Č Ć Đ É Ě ě Í Î Ĺ Ľ Ń Ô Ö Ř Š ſ Ţ Ú Û Ů Ž, występujące w zweryfikowanych, prawidłowych postaciach imion w innych językach. Po dokonaniu eliminacji wszystkich wątpliwych form imion liczba okazów wyniosła 48 517 037, natomiast liczba typów imion 37 089. Poniżej przedstawiono krótki opis błędów występujących w materiale badawczym oraz uzasadnienie eliminacji zawierających je rekordów.

W rubryce „imię pierwsze” ujawniło się wiele formuł w postaci zestawienia więcej niż jednego imienia, a czasem także innych niż imię określeń lub elementów, oddzielonych spacją (np. ANNA MARIA, IRENA MARIA, JAN JÓZEF, KAZIMIERZ JAN) lub dywizem (np.

ANNA-LIZA, MARIA-MAGDALENA, JEAN-PAUL, KAZIMIERZ-WŁADYSŁAW). Poza formułami zawierającymi dwa, a nawet trzy lub cztery imiona (jak np. *MARIA JOLANTA KRYSTYNA, ALBA VIVIANNA IWONA, THO PHUONG ANH, KIM ANNA MAGDALENA MARIA*), ujawniły się także formuły, gdzie spacja pełniła funkcję delimitacyjną nie dla imion, ale też np. spójnika *vel*, stosowanego dla połączenia określeń równoważnych czy synonimicznych, jak w przypadku formuł (*TEA VEL LEA, AGAFIA VEL AGATA*, liczebników porządkowych (*JÓZEF I, KAROL II, JAN III*) czy określeń typu *junior* (*JACEK JR, SŁAWOMIR JUNIOR*). Marginalnie pojawiły się też imiona zestawione z hiszpańskojęzycznymi wyrażeniami przyimkowymi określającymi pochodzenie (*BARBARA DE LOS ANGELES, YAREMI DE LAS MERCEDES*). Wystąpiły też przypadki, gdy drugi element był zniekształcony (niekompletny), jak np. *BRONISŁAWA L, ZBIGNIEW STA, ZYTA MA, JAN-*). Większość wymienionych wyżej formuł imienniczych nie spełnia kryteriów bycia imieniem pierwszym rozumianym jako pojedyncze imię. Teoretycznie w celu skorygowania takich formuł moglibyśmy po prostu „obciąć” wszystkie elementy poza imieniem, które występuje jako pierwsze w szeregu, pojawia się jednak wątpliwość dotycząca np. imion w postaci transliterowanej z języków afrykańskich (np. *A-NSOMPU*) czy azjatyckich (np. *AN-DONG*). Podobnie trudno rozsądzić w każdym przypadku użycia formuły z dywizem, kiedy stosowano imię złożone, a kiedy użycie dywizu stanowiło tylko pewną konwencję lub błąd osoby, która wprowadzała dane do bazy PESEL. Ponadto użycie spacji nie zawsze wiązało się z imieniem w postaci wieloelementowego szeregu, odnotowano bowiem także wypadki, gdy doszło do przypadkowego podziału imienia (np. *M IECZYŚLAW, JA NINA, ŁU UKASZ*) lub też spacja wystąpiła w miejscu brakującego znaku, zwykle diakrytycznego (np. *CZES AW , STANIS AW*). O ile takie błędy można by wyeliminować, to jednak nie jest to już możliwe w przypadku takich formacji jak *SAM_ON*, gdzie nie możemy rozstrzygnąć, czy chodzi o imię *Samon, Samson* czy też może szereg *Sam On* (występujący m.in. w kambodżańskiej antroponimii).

W całym pierwotnym materiale odnotowano też wiele typów, które zawierały znaki inne niż spacja, dywiz, takie jak *, . ; & %* itd. Część z nich miało defekty wynikające z błędnego użycia znaku, np. *KAZIM9ERZ, STANÍŚLAWA, WŁADYßŁAW*, nadwyżkowego użycia znaku (kropki, cyfry i in.), np. *JADWIGA., OL.IWIA, TADEUSZ.1*) lub przeniesienia treści z innego pola bazy danych (np. *EWA.NP=DEC*). Odnotowano też przypadki zastąpienia litery odpowiadającym jej punktem kodowym w standardzie Unicode, np. *L<U+03AB>DIA* (właśc. *LÝDIA*), *VI<U+IEC6>T* (właśc. *VIỆT*) itd. O ile te przypadki dałoby się skorygować, to nie możliwe było to w przypadku form takich jak *NICOL<U+01FE>* (co oddaje nigdzie

niespotykaną postać *NICOLŌ*), nie wiemy bowiem jaki znak powinien znaleźć się w miejscu błędnego *Ō*. Wyeliminowano również wszystkie inne pozycje, które zawierały kropkę użytą w skrócie imienia, np. w formacjach *J.*, *M.*, *MD.*, *MHD.*, *Y.*, gdzie na podstawie skrótu niemożliwe jest odtworzenie pełnej postaci pierwszego imienia.³

Konieczne było też wyeliminowanie postaci zawierających znaki o niejednoznacznej roli. O ile w formule *MATTHA%US* ciąg *A%* odpowiada zapewne literze *Ä*, to nie możemy ze stuprocentową pewnością stwierdzić, że w przykładach: *KA%TE*, *CA%CILIE*, *BE%RBEL*, *ZOE%* itd. znak *%* zawsze odpowiadał umlautowi. To samo dotyczy apostrofu (*'*). W formule typu *O'NEIL* znak jest ten użyty prawidłowo, jednak nie można jednoznacznie zinterpretować jego roli w innych imionach, gdyż raz może odpowiadać ukośnej kresce (akutowi) nad literą następującą, np. *NICOL'E* (popr. *NICOLÉ*), kiedy indziej pełnić funkcję miękkości, jak w transliterowanych imionach rosyjskich (np. *OL'GA*) lub mieć inną, trudną do zinterpretowania rolę (*ANN'ESSY*, *MARCO'*, *A'ISHAT*, *JAKUB'*).

Inną problematyczną kwestią było współwystępowanie w bazie błędnych wariantów graficznych tego samego imienia powstałych podczas wprowadzania danych do bazy PESEL. Przykładem może być niemieckie imię *Günter*. Występujący w nim grafem *ü* został zapisany na kilka sposobów: *GÜNTER*, *GÚNTER*, *GŮENTER*, *G;UNTER*, *GU'NTER*, *GUNTER*. Każda z tych postaci została zapisana w innej, błędnej manierze graficznej (poza *GÜNTER*), co powoduje, że komputer rozpoznaje je jako odrębne formy. Teoretycznie można by sprowadzić je do formy poprawnej *GÜNTER*, jednak prześledzenie całej bazy pod kątem występowania takich potencjalnych błędnych wariantów graficznych innych imion jest zadaniem trudnym do zrealizowania dla jednej osoby, jeśli w ogóle jest to możliwe (zwłaszcza, że w materiale występuje wiele imion genetycznie niepolskich, a nawet pozaeuropejskich).

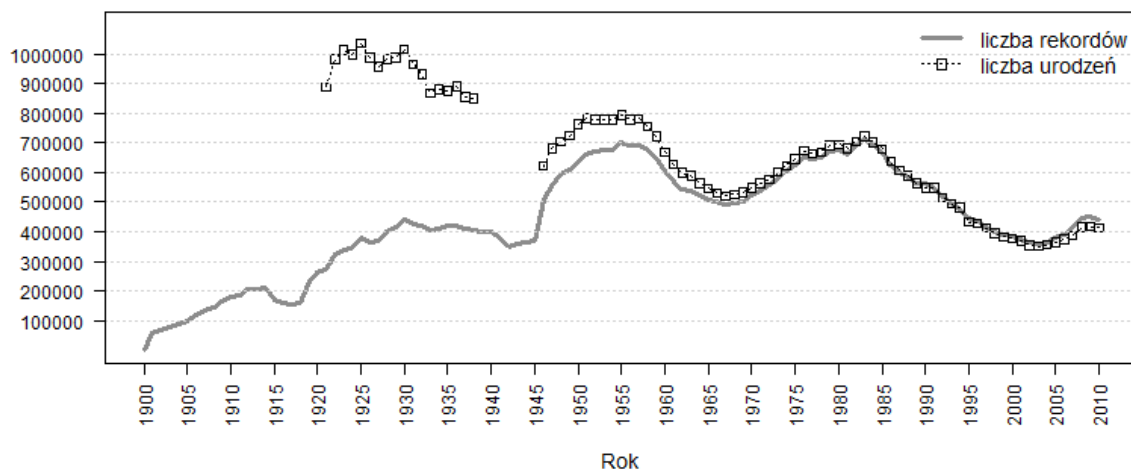
Mając świadomość, że wyeliminowanie wszystkich wątpliwych form jest bardzo trudne, uznano za wyjście kompromisowe maksymalnie ograniczyć ich liczbę przez eliminację stosunkowo nieznacznej liczby okazów zawierających błędne znaki diakrytyczne oraz znaki nieliterowe. Doświadczenie innych badaczy wskazuje też, że eliminacja niewielkiej części materiału z tak dużej bazy danych, jaką stanowi baza PESEL, jest wyjściem rozsądniejszym

³ Skróty *MHD.*, *MD.* mogą oddawać wiele postaci: *Mahmud*, *Muhammad*, *Mohammed* itd.

niż usilne próby jego pełnego skorygowania, które mogą przynieść skutki odwrotne od zamierzonych.⁴

Na wykresie z rys. 1 porównano liczbę rekordów (wielkość korpusów imienniczych) w materiale badawczym po jego oczyszczeniu z dostępnymi danymi dotyczącymi liczby urodzeń w Polsce w poszczególnych latach.

Rys. 1. Porównanie liczby rekordów w materiale badawczym oraz liczby urodzeń w kolejnych latach

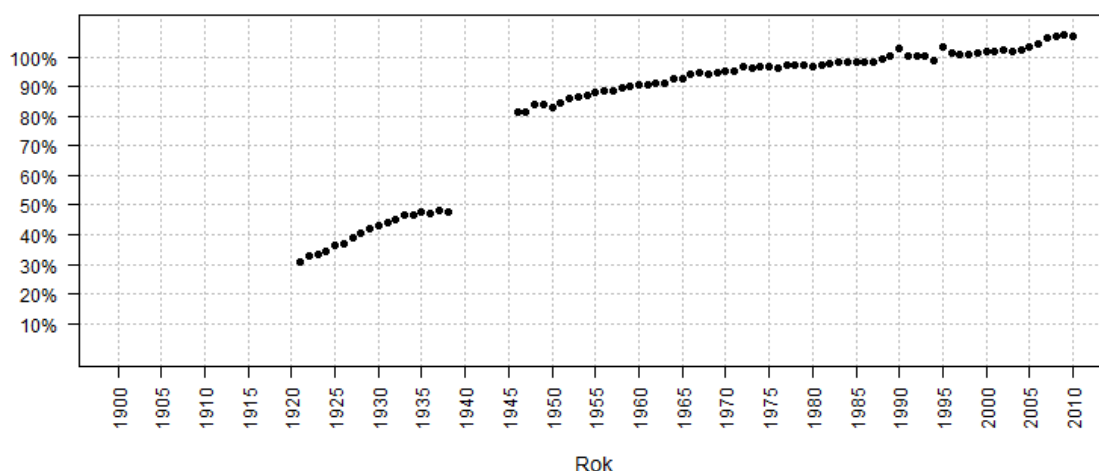


Wspomniano już, że materiał z PESEL uznaje się za w pełni reprezentatywny dla imion nadawanych od 1984 r., zaś w przypadku wcześniejszych lat możemy mówić o mniejszych lub większych próbach z populacji, ograniczonych do osób, które żyły w roku 1984, a zatem ich liczebność zależna jest od czynników biologicznych czy demograficznych. Z danych przedstawionych na powyższym wykresie wynika, że liczba osób ujęta w naszym materiale badawczym w bardzo dużym stopniu pokrywa rzeczywistą liczbę osób urodzonych także we wcześniejszych latach, począwszy od zakończenia II wojny światowej. Na kolejnym wykresie (rys. 2) przedstawiono procentowy stosunek liczby rekordów w naszym materiale do rzeczywistej liczby urodzeń odnotowanych w Polsce. W roku 1946 wyniósł on ponad 81% i systematycznie rósł, by już od lat 60. osiągać poziom ponad 90%, a w ostatnich latach nawet powyżej 100% (kwestia ta zostanie omówiona w dalszej części tekstu).

Jeśli chodzi o porównanie z danymi dotyczącymi liczby urodzeń dostępnymi dla lat 1921–1938, to stosunek ten już jest znacznie mniejszy i wynosi od niecałych 31% w roku 1921 do ok. 48% pod koniec tego okresu. Należy zakładać, że z każdym kolejnym rokiem wstecz

⁴ D. K. Tucker, rozważając zasadność korekt „podejrzanych” imion i nazwisk w elektronicznej bazie danych powstałej na podstawie spisu powszechnego z 1881 w Anglii i Walii, stwierdził ostatecznie: „It is tempting to attempt correction but the danger is that a correct spelling could be changed and that an incorrect spelling could be changed to a correct spelling but not the one for the name of the enumerated person” (Tucker 2003: 214).

Rys. 2. Stosunek liczby rekordów w materiale badawczym do rzeczywistej liczby urodzeń odnotowanych w Polsce



procent ten będzie systematycznie malał. Mając na uwadze tak duże dysproporcje liczbowe między liczbą rekordów w poszczególnych latach, w części pracy poświęconej analizie wskaźników bogactwa leksykalnego, która wymaga operowania porównywalnej wielkości korpusami imiennicznymi, konieczne było sporządzenie mniejszych próbek losowych dla poszczególnych lat (szczegółowo zostało to opisane w części III.1 i III.2).

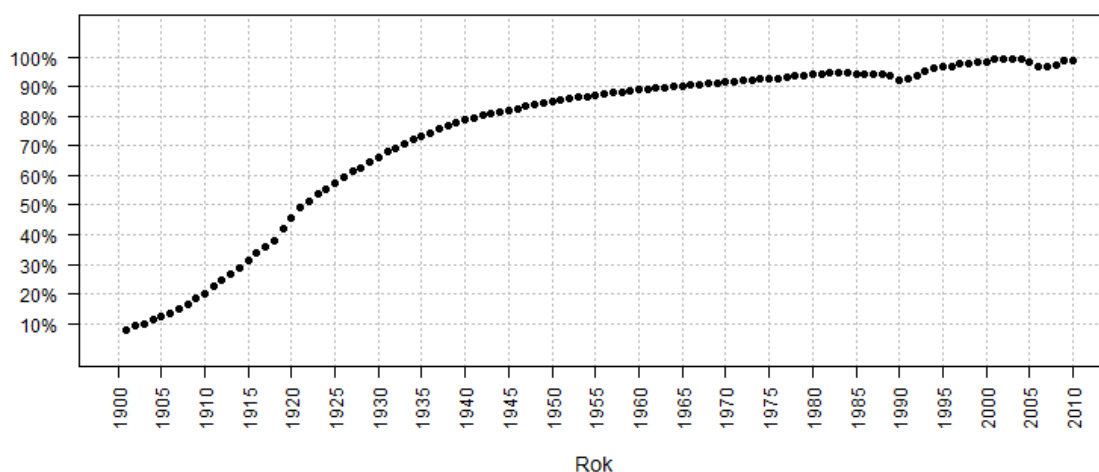
Warto w tym miejscu poczynić kilka uwag dotyczących deficytów w materiale badawczym związanych z uwarunkowaniami historyczno-demograficznymi. Widoczne na wykresie z rys. 1 dwa wyraźne spadki liczby rekordów z okresu I i II wojny światowej wiążą się z wysoką śmiertelnością w tym okresie, co przełożyło się na stosunkowo niską reprezentację w materiale badawczym. Jeśli chodzi o drugą wojnę światową, to jej skutkiem był także inny poważny deficyt, mianowicie niewątpliwe niedoreprezentowanie imion polskich obywateli pochodzenia żydowskiego urodzonych zarówno w trakcie, jak i na długo przed II wojną światową. W wyniku Holokaustu według różnych szacunków życie straciło od 2,7 do 3 mln Żydów polskich, co stanowiło od 83,1% do 92,3% ich przedwojennej populacji, liczącej w roku 1933 ok. 3,25 mln (Niewyk 2000: 421). Musimy sobie także zdawać sprawę z faktu, iż w materiale badawczym nie znalazło się wiele informacji o osobach, które w wyniku przesunięcia granic po 1945 r. utraciły obywatelstwo polskie, co dotyczy zarówno Polaków, którzy po wojnie pozostali poza granicami kraju, a także znacznej części przedwojennych obywateli polskich pochodzenia ukraińskiego, białoruskiego czy litewskiego, ale też niemieckiego.

Inną sprawą wymagającą komentarza jest z kolei zjawisko obserwowane od roku 1990, polegające na występowaniu w naszym materiale wyższej liczby osób, którym nadano imię niż osób urodzonych. Wiąże się ono zapewne z faktem, iż dane o urodzeniach pozyskane

z GUS dotyczą jedynie obszaru Polski, natomiast nie uwzględniają obywateli Polski urodzonych poza granicami kraju. Zjawisko to nasiliło się zwłaszcza w ostatnich latach badanego okresu i wynika ze wzmożonego ruchu emigracyjnego.

Innym elementem materiału badawczego, który również wymagał weryfikacji i korekty, były kody terytorialne TERC opisujące miejsce rejestracji urodzenia danej osoby. Jak już nadmieniono wcześniej, kody TERC nie są przypisane w pozyskanym z PESEL materiale do wszystkich osób. Dotyczy to zwłaszcza danych o osobach urodzonych przed 1945 r. (zwykle w związku z faktem urodzenia poza powojennymi granicami Polski), ale także później. Na wykresie z rys. 3 przedstawiono procentowy udział rekordów zawierających kody TERC.

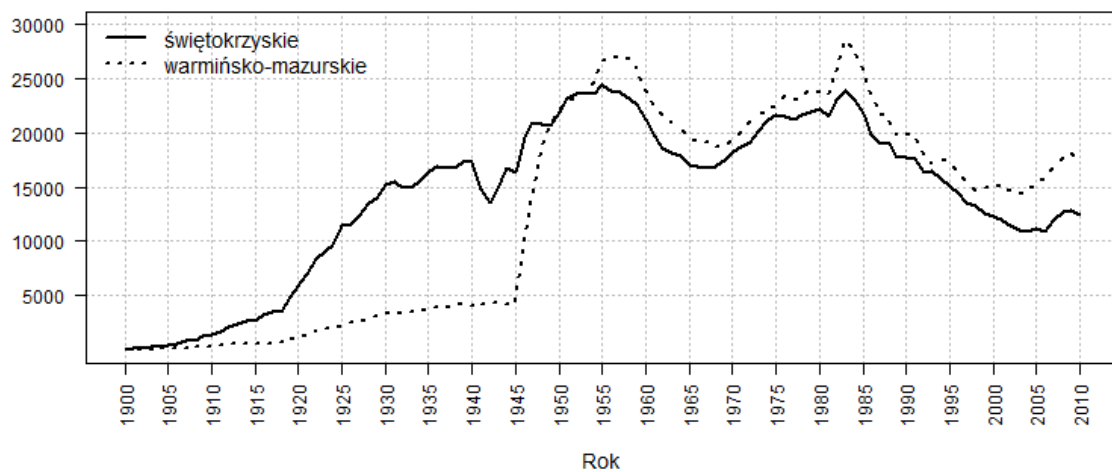
Rys. 3. Procentowy udział w materiale badawczym z kolejnych lat rekordów zawierających identyfikatory TERC



Widzimy, że w przypadku rekordów dotyczących osób urodzonych w pierwszych dwu dekadach w rozpatrywanym okresie udział ten był bardzo niski. Wiąże się to z faktem, że miejsce urodzenia wielu z tych osób znajduje się poza obecnymi granicami Polski, co uniemożliwiało przypisanie im kodu odnoszącego się do współczesnego podziału administracyjnego. Musimy mieć też na uwadze kwestię, że nawet w przypadku późniejszych lat do roku 1945 r. gdy udział ten znacznie się zwiększył, informacje o miejscu urodzenia dotyczą w ogromnej mierze obszarów II RP, które po tym roku pozostały w granicach Polski. W znikomym zaś stopniu występują rekordy zawierające informacje o urodzeniach na obszarach, które przed 1945 r. nie należały terytorium Polski (dotyczy to przede wszystkim województw dolnośląskiego, lubuskiego, opolskiego, zachodniopomorskiego czy warmińsko-mazurskiego), co wynika z oczywistego faktu, że były to tereny słabo wówczas zaludnione przez powojennych obywateli Polski (ogromna większość z ich ówczesnych mieszkańców

została po 1945 r. wysiedlona). Zjawisko to dobrze ilustruje wykres na rys. 4, gdzie porównano liczbę rekordów zawierających identyfikator TERC dla obszaru dwóch współczesnych województw podobnych obecnie pod względem liczby ludności – świętokrzyskiego (przed 1939 r. należącego w całości do II RP) i warmińsko-mazurskiego (niemal w całości należącego przed 1939 r. do Niemiec). Łatwo zauważyć, że w przypadku tego drugiego do 1945 r. dysponujemy szczątkowymi informacjami i dopiero później, po osiedleniu się w latach powojennych ludności polskiej, możemy mówić o w miarę pełnych danych dotyczących miejsca rejestracji urodzenia na jego obszarze.

Rys. 4. Porównanie liczby rekordów zawierających identyfikator TERC dla współczesnych województw świętokrzyskiego i warmińsko-mazurskiego



W związku z zarysowanymi powyżej kwestiami niektóre analizy porównawcze dotyczące zróżnicowania geograficznego imiennictwa mogły być z oczywistych względów prowadzone tylko w odniesieniu do okresu po II wojnie światowej (piszę o tym także w kontekście doboru materiału w rozdziale III.2).

Niektóre rekordy w bazie PESEL zawierały też przestarzałe kody TERC, które uległy w ostatnich latach zmianie, co uniemożliwiało przypisanie na ich podstawie poprawnej lokalizacji geograficznej niektórym osobom. Wszystkie kody zidentyfikowane jako przestarzałe (264) zostały dostosowane do stanu współczesnego.

I.3.2. Imiona w bazie PESEL na tle innych rejestrów w dotychczasowych badaniach imiennictwa

W zależności od zakresu i skali badania oraz ujęcia metodologicznego w pracach dotyczących imiennictwa⁵ stosuje się różnego rodzaju materiał badawczy, ale też sam charakter i zakres, a przede wszystkim dostępność tego materiału wyznaczają możliwości jego analizy, a także wpływają na miarodajność płynących z niej wniosków. Jak mogliśmy się przekonać w poprzedniej części, dostępne dane z bazy PESEL nie są „idealne”, zwłaszcza jeśli chodzi o okres przed 1945 r., warto jednak przyjrzeć się, jak prezentują się na tle innych rejestrów, wykorzystywanych w badaniach współczesnego imiennictwa w Polsce i innych krajach oraz jakie są możliwości ich wykorzystania do szeroko zakrojonych analiz porównawczych, będących celem niniejszej pracy.

Niemieccy badacze D. Huschka i G. Wagner (2010) w pracy poświęconej problemom związanym ze stosowaniem metod statystycznych w badaniach nad imiennictwem zwracają uwagę na bardzo istotną kwestię reprezentatywności materiału źródłowego w kontekście rozmaitych porównań (np. w zakresie imiennictwa różnych państw, regionów, grup społecznych, pokoleń itd.). W dostępnych pracach antroponomastycznych, zawierających elementy analizy statystycznej czy socjolingwistycznej, najczęściej spotykane są następujące typy źródeł informacji o imiennictwie: książki telefoniczne (w polskich pracach raczej nie wykorzystywane), księgi metrykalne gromadzone w parafiach lub urzędach stanu cywilnego, ankiety badawcze, bazy danych oficjalnych rejestrów państwowych oraz wielkoskalowe badania sondażowe.

Książki telefoniczne, którymi w swoich pracach posługiwali się m.in. Eshel (2001) czy Tucker (2001), stanowią obszerny, łatwo dostępny, ale – jak słusznie zauważają Huschka i Wagner (2010: 5–6) – jednocześnie bardzo niereprezentatywny materiał badawczy, gdyż pokrywają przeważnie stosunkowo wąską część struktury społeczeństwa. Przede wszystkim na samą obecność w książce telefonicznej lub jej brak miały wpływ mają czynniki społeczne – mogą one pomijać nie tylko osoby, które z powodu niskiego statusu ekonomicznego nie posiadały telefonu, ale też przedstawicieli tzw. wyższych sfer, którzy z powodu ochrony prywatności nie upubliczniali swoich danych. Ponadto w rejestrach abonentów telefonicznych niedoreprezentowane są kobiety (zwłaszcza zamężne) oraz dzieci, gdyż zwykle w danym

⁵ Nie będziemy w tym miejscu skupiać się na opracowaniach syntetycznych w rodzaju *Słownika staropolskich nazw osobowych*, które w zakresie materiałów są opracowaniami wtórnymi. Chodzi przede wszystkim o takie prace, których autorzy bazują na tzw. surowym, nieopracowanym lub tylko częściowo opracowanym materiale.

rekordzie pojawia się imię i nazwisko męża czy ojca. Niemożliwe jest też na podstawie takiego źródła określenie momentu nadania imienia, ani też jego miejsca (dysponujemy bowiem jedynie informacją o miejscu zamieszkania). Mamy zatem w przypadku książki telefonicznej do czynienia z wycinkowym zbiorem imion używanych (a nie nadanych) w momencie jej ukazania.

Księgi metrykalne gromadzone przez urzędy stanu cywilnego czy parafie, stanowią zdecydowanie pełniejsze i wiarygodniejsze źródło wiedzy o imiennictwie danego obszaru, gdyż zwykle obejmują wszystkie osoby na nim urodzone (w przypadku ksiąg parafialnych mowa oczywiście o członkach danej wspólnoty wyznaniowej). Zawierają ponadto szereg informacji dodatkowych, ukazujących kontekst społeczny i socjalny towarzyszący nadawaniu imienia, takich jak dokładna data i miejsce urodzenia, imiona rodziców, często też informacje o ich wykształceniu, zawodzie czy pochodzeniu społecznym, co umożliwia dokonanie pogłębionej analizy socjolingwistycznej. Pozyskanie takiego materiału wiąże się jednak ze żmudną kwerendą źródeł, toteż bazujące na nich prace, dominujące wśród polskich tekstów⁶, mają zakres lokalny – autorzy zajmują się w nich zwykle imionami nadawanymi w pojedynczych miejscowościach lub parafiach, w jednym lub kilku przekrojach czasowych. To samo dotyczy prac, do których materiał badawczy pozyskano metodą wywiadów ankietowych przeprowadzanych wśród rodziców nowo narodzonych dzieci m.in. w szpitalach położniczych⁷. Tu, rzecz jasna, nie możemy mówić o odzwierciedleniu pełnego zasobu imienniczego w badanej miejscowości, a jedynie o niewielkich próbach z populacji, natomiast podobnie jak w przypadku materiału pochodzącego z ksiąg metrykalnych, ankiety uwzględniają szereg czynników pozajęzykowych towarzyszących nadaniu imienia. Przeprowadzenie i analiza ankiet przez jednego badacza możliwe są zwykle tylko w obrębie niewielkiego obszaru i w ograniczonym zasięgu czasowym. Choć prac bazujących na materiale pozyskanym z ksiąg metrykalnych lub drogą ankietową w ostatnich kilku dziesięcioleciach znacznie przybyło i stanowią one bardzo cenne źródło wiedzy o funkcjonowaniu lokalnych systemów imiennicznych, to nadal aktualna wydaje się być opinia H. Borka sprzed przeszło 40 lat, podsumowująca stan badań nad współczesnym polskim imiennictwem: „W sumie jednak tego niewiele, a przede wszystkim ma charakter dość

⁶ Najczęściej źródło materiału badawczego stanowią w nich księgi parafialne (np. Umińska-Tytoń 1988, Plichta 1998, Tomecka 1994), zwłaszcza jeśli chodzi o starsze warstwy chronologiczne, rzadziej zaś pojawiają się wypisy z ksiąg urzędów stanu cywilnego (Gajda 1973, Umińska-Tytoń 1983, Abramowicz 1993, Rudnicka-Fira 2011, Iglewska 2015). Można spotkać się też z pracami, które stosują bardziej zdywersyfikowane źródła, łącząc księgi parafialne, akta miejskie, dokumenty cechów itd. (np. Magda-Czekaj 2000).

⁷ Por. np. Folfasiński 1947–1949, Borek 1978, Breza 1987, Lewaszkiewicz, Rzepka 1988.

przypadkowy i rozproszony, częściowo tylko podporządkowany określonym dyrektywom metodologicznym” (Borek 1978: 166). Brak jest bowiem skoordynowanych i regularnie prowadzonych lokalnych badań nad imiennictwem, których wyniki dałoby się usystematyzować i ukazać jego czasowe i regionalne zróżnicowanie oraz długofalowe tendencje w tym zakresie w odniesieniu do całego obszaru kraju w dłuższej perspektywie czasowej.

Szansę na zrealizowanie takiego celu stwarza wykorzystanie jako materiału badawczego państwowych rejestrów ludności w postaci elektronicznych baz danych. Taką sytuację Huschka i Gerhards (2010: 6) określają mianem „złotego standardu” w badaniach socjonomastycznych, gdyż rejestry takie z założenia powinny obejmować wszystkich obywateli danego kraju i zawierać szczegółowe informacje dotyczące ich czasu i miejsca urodzenia, pochodzenia, rodziców itd. W praktyce pełny dostęp do takich baz danych jest rzadko możliwy z uwagi na restrykcyjne przepisy dotyczące ochrony danych osobowych. W ostatnich latach – zarówno w Polsce, jak i w innych krajach, nastąpił jednak znaczny progres w zakresie możliwości częściowego dostępu do wielkoskalowych danych zawartych w takich bazach danych oraz ich udostępniania.

Jednym z takich rejestrów jest baza PESEL, stanowiąca źródło materiału badawczego dla tej pracy. Publiczny dostęp do podstawowych danych o imionach w niej zawartych pojawił się dopiero niedawno⁸, w nieco szerszym, acz również ograniczonym stopniu udostępniane są na potrzeby badań naukowych. Zakres informacji udostępnianych przez administratora PESEL został omówiony już wcześniej. Należy jednak w tym miejscu zaznaczyć, że baza ta była wykorzystywana w polskich badaniach nad imiennictwem dużo wcześniej. W 1995 r. K. Rymut opublikował *Słownik imion współcześnie w Polsce używanych*, który bazował na informacjach z PESEL zgromadzonych tam do 1994 r. Jest to słownik frekwencyjny, który przedstawia ogólną liczbę obywateli polskich noszących dane każde imię (także w podziale na dekady), płeć oraz miejsce zamieszkania na poziomie ówczesnych województw, wraz z podaniem liczby nosicieli imienia w danym województwie. Sam słownik, stanowiąc najpełniejszy ogólnie dostępny rejestr imion, stał się źródłem materiału badawczego dla innych prac, które koncentrowały się na zmienności polskiego systemu imienniczego w czasie czy jego właściwości statystycznych (por. np. Nowik 1998, Skowronek 2001, Malec 2007).

⁸ Na stronie Ministerstwa Cyfryzacji od końca 2017 r. zaczęto publikować dane o imionach osób (mających więcej niż 1 nosiciela) nadawanych w skali ogólnopolskiej oraz poszczególnych województwach. Dane te obejmują lata 2015–2019. Od 2019 r. publikowane są też informacje o liczbie nosicieli pierwszych i drugich imion (przedstawiono je jednak oddzielnie!), dotyczące osób żyjących, urodzonych przed 1 stycznia 2019 r. (<https://www.gov.pl/web/cyfryzacja/dodatkowe-materialy>, [dostęp: 1 marca 2019 r.]).

Abstrahując od wspomnianych wcześniej braków w systemie PESEL, mających swe obiektywne przyczyny, trzeba zwrócić uwagę na pewne mankamenty słownika K. Rymuta traktowanego jako źródło materiału. Autor z niejasnych powodów usunął z niego „imiona typowe dla antroponimii Dalekiego Wschodu, np. wietnamskie, chińskie, japońskie”, pozostawiając w nim jednak „imiona związane z religią mahometańską, z ludami arabskimi i tureckimi” (Rymut 1995: VIII), zacierając w ten sposób obraz imiennictwa polskiego XX w. i wykluczając pośrednio z grona obywateli polskich część osób pochodzenia azjatyckiego. Dane o rozmieszczeniu geograficznym nosicieli imion nie są schronologizowane i dotyczą tylko miejsca zamieszkania danej osoby, które nie musi być tożsame z miejscem urodzenia – pozbawia nas to możliwości prześledzenia wpływu miejsca pochodzenia osób dokonujących aktu nominacji na kształt jego antroponimii w określonym czasie. Wreszcie – choć trudno czynić autorowi zarzut z tego powodu – z uwagi na tradycyjną, „papierową” formę wydania słownika bardzo utrudnione jest wykorzystanie go na szerszą skalę w badaniach dotyczących całości polskiego imiennictwa, gdyż wymagałoby to niejako powtórnego wprowadzenia do komputera danych w nim zawartych.

Poza słownikiem K. Rymuta dane z bazy PESEL posłużyły też jako źródło materiału badawczego dla cyklu artykułów poświęconych przemianom ilościowo-jakościowym w polskim imiennictwie w okresie 1995–2010 (Cieślikowa 2013, Czopek-Kopciuch 2013, Skowronek 2013, Swoboda 2013) oraz w tekście dotyczącym jego regionalnego zróżnicowania ilościowego w latach 1951–2010 (Swoboda 2015). Ponadto od 2013 r. Ministerstwo Cyfryzacji publikuje na podstawie danych z PESEL statystyki dotyczące frekwencji imion nadanych w całym kraju oraz w poszczególnych województwach, jednak z uwagi na wrażliwość danych nie obejmują one imion o frekwencji mniejszej niż 2 w danym roku.

Rejestry ludności podobne do bazy PESEL funkcjonują także w innych krajach. Z uwagi na częste wykorzystywanie w badaniach antroponomastycznych na uwagę zasługują zwłaszcza dwa z nich – amerykański, prowadzony w ramach agencji rządowej United States Social Security Administration (SSA), oraz holenderski, Basisregistratie Personen (BRP), będący odpowiednikiem polskiego PESEL-u.

Amerykańskie SSA na swoich stronach internetowych prowadzi serwis statystyczny specjalnie poświęcony imionom nadawanym dzieciom w USA.⁹ Pełni on przede wszystkim rolę popularyzatorską i stanowi źródło referencji dla osób planujących nadanie dziecku

⁹ <https://www.ssa.gov/oact/babynames/> [dostęp: 28.12.2018]

imienia. Można w nim prześledzić listy najpopularniejszych imion w całym kraju i w poszczególnych stanach, zmienność popularności imienia w czasie itd. Z punktu widzenia badaczy, zainteresowanych bardziej szczegółowymi analizami, najistotniejsza jest możliwość pobrania danych o frekwencji imion na poziomie poszczególnych stanów w kolejnych latach od 1880 r. po dzień dzisiejszy. Rejestr nie obejmuje jednak imion najrzadszych, które zostały nadane mniej niż 5 razy w danym stanie. Ponadto, choć obecnie obejmuje niemal wszystkie osoby urodzone w USA, to jednak nie pokrywa całej populacji osób urodzonych w poszczególnych latach, co dotyczy zwłaszcza okresu sprzed powstania rejestru SSA w 1937 r.¹⁰ – w pierwszych kilku dekadach pokrycie to wynosi nie więcej niż 20%. Niemniej jednak ciągle jest to największa publicznie dostępna baza informacji o imiennictwie. Otwarty dostęp do danych zawartych w SSA przyciąga zainteresowanie badaczy z całego świata, co zaowocowało w ostatnich latach powstaniem licznych prac o charakterze statystycznym poświęconych czasowemu i przestrzennemu zróżnicowaniu imiennictwa w USA (por. np. Evans 2007, Tucker 2009, Twenge i in. 2010, 2016, Li 2012, Barucca i in. 2015, Pomorski i in. 2016).

Drugim z omawianych rejestrów jest holenderski Basisregistratie Personen (BRP)¹¹, zawierający dane o mieszkańcach Holandii urodzonych od 1880 r. do dzisiaj. Rejestr ten niemal w pełni pokrywa całą populację od roku 1930, zaś w przypadku osób urodzonych przed tą datą dysponujemy większymi lub mniejszymi próbami – dla roku 1880 dysponujemy informacjami o ok. 30% osób (Bloothoof, Mandemakers 2014: 272). Na jego kanwie opracowano popularny korpus imion używanych w Holandii – Nederlandse Voornamenbank¹², który umożliwia dostęp do informacji o frekwencji rocznej ponad 600 tys. imion używanych w Holandii w całym okresie, a także o rozmieszczeniu geograficznym nosicieli imion w 2014 r. do poziomu gmin – dane te prezentowane są w formie automatycznie generowanych wykresów i kartogramów. Co więcej, możliwe są też analizy pod tym kątem całych grup imion odznaczających się podobnym brzmieniem lub też pochodzeniem językowym (duża część imion opatrzona jest informacjami etymologicznymi). Dostęp do bardziej szczegółowych danych o wszystkich osobach ujętych w BRP jest możliwy tylko w przypadku badań naukowych, jednak wówczas zakres tych danych jest dużo szerszy

¹⁰ Dane o imionach w SSA bazują bowiem na wnioskach o wydanie kart ubezpieczenia społecznego. Rejestr ten został wdrożony w 1937 r. Spośród osób urodzonych wcześniej znalazły się w nim tylko te osoby, które wystąpiły o nadanie numeru SSA. W kolejnych latach liczba osób objęta systemem ubezpieczeń społecznych systematycznie rosła i obejmowała niemal wszystkie osoby urodzone w USA.

¹¹ <https://www.rvig.nl/brp/brp-als-basisregistratie/> [dostęp: 28.12.2018]

¹² <http://www.meertens.knaw.nl/nvb/> [dostęp: 28.12.2018]

niż ma to miejsce w przypadku PESEL czy SSA. Badacze dysponują bowiem dodatkowymi informacjami takimi jak drugie imię, nazwisko, dokładna data i miejsce urodzenia, miejsce zamieszkania osób ujętych w rejestrze, a poszczególne rekordy powiązane są także z informacjami o rodzicach. Stwarza to możliwość przeprowadzenia wszechstronnych i pogłębionych analiz statystycznych, lingwistycznych i socjoonomastycznych, których efekty możemy śledzić zwłaszcza w pracach opublikowanych przez holenderskiego badacza G. Bloothoofa i współpracowników (por. np. Bloothoof, Groot 2008, Bloothoof, Schraagen 2011, Bloothoof, Mandemakers 2014, Bloothoof, Onland 2016).

Materiał badawczy pozyskany na potrzeby niniejszej pracy z bazy PESEL jest pod względem zasobności i ilości informacji porównywalny z rejestrami przedstawionymi powyżej. Stanowi bogate, choć niepełne źródło informacji na temat imiennictwa Polaków w XX i XXI w. Jak mogliśmy się przekonać, nie istnieje właściwie żaden rejestr (a przynajmniej żaden dostępny publicznie lub chociaż na potrzeby badawcze), który pokrywałby w 100% daną populację, zwłaszcza w przypadku lat przedwojennych. Niemniej ilość danych w PESEL odnoszących się do dawniejszych z lat rozpatrywanego okresu wydaje się stanowić reprezentatywną próbę dla starszego imiennictwa. Przy zachowaniu odpowiednich procedur (zob. rozdziały III.1, III.2) może stanowić źródło porównań dla lat późniejszych. Spostrzeżenia te dotyczą jednak imiennictwa ogólnopolskiego, bez uwzględnienia jego dystrybucji przestrzennej. Chcąc dokonywać regionalnych porównań antroponimii, możemy tylko w ograniczonym stopniu sięgać do okresu sprzed 1945 r., gdyż identyfikatory przynależności terytorialnej poszczególnych osób dotyczą w głównej mierze obszarów, które przed II wojną światową znajdowały się na terenie Polski. Mając na uwadze te ograniczenia, można założyć, że materiał z bazy PESEL pozwala na przeprowadzenie zakładanych analiz zróżnicowania imiennictwa z uwzględnieniem takich czynników jak przynależność czasowa, terytorialna oraz płeć osób otrzymujących imiona.

II. Zagadnienia teoretyczno-metodologiczne

II.1. Stan badań

II.1.1. Zarys stanu badań nad polskim imiennictwem

Badania antroponimiczne mają w Polsce długą tradycję, a problematyka samego imiennictwa doczekała się bogatej literatury. Syntezę dotychczasowych badań nad imionami stanowi książka M. Malec *Imię w polskiej antroponimii i kulturze* (2001), w której autorka wszechstronnie opisała najrozmaitsze aspekty funkcjonowania imion w języku, kulturze czy komunikacji społecznej, a także przedstawiła główne nurty w badaniach nad imiennictwem. Znajdziemy w niej również szeroki przegląd literatury poświęconej sprawom imiennictwa. W tej części pracy pokrótce zarysowany zostanie stan badań nad imionami osób żyjących w Polsce obecnie lub przeszłości, pominię natomiast zagadnienia imiennictwa w onomastyce literackiej, folklorze itp.

Pierwsze naukowe zainteresowania imionami skupiały się na okresie staropolskim i można powiedzieć, że jest to warstwa kompleksowo już opisana. Za pierwszą, *stricte* naukową pracę poświęconą imionom uznaje się opracowanie Witolda Taszyckiego *Najdawniejsze polskie imiona osobowe* (1925), której celem był opis budowy słowotwórczej oraz klasyfikacja semantyczno-strukturalna imion słowiańskich, odnotowanych w źródłach historycznych do roku 1300. Ujęcie zastosowane w tej i innych zainicjowanych przez Taszyckiego pracach na długo wyznaczyło główny nurt opracowań antroponomastycznych, w których przodował paradygmat strukturalistyczny, skupiony głównie na budowie słowotwórczej i etymologii najstarszych antroponimów. Sztandarowym dziełem polskiej leksykografii antroponomastycznej dotyczącej okresu staropolskiego jest siedmiotomowy *Słownik staropolskich nazw osobowych* (Taszycki (red.), 1965–1987), dokumentujący antroponimy odnotowane w źródłach historycznych z okresu X–XV w., którego rozwinięcie stanowi *Słownik etymologiczno-motywacyjny staropolskich nazw osobowych* (Cieślíkowa i in. (red.) 1995–2002), w którego poszczególnych tomach objaśniono staropolskie antroponimy w podziale na różne grupy motywacyjne (odapelatywne, zapożyczone – chrześcijańskie i innojęzyczne, odmiejscowe i odetniczne). Na podstawie materiału z SSNO oraz jego kartoteki powstały liczne monografie i artykuły dotyczące w głównej mierze słowotwórstwa antroponimów rodzimych oraz pochodzenia i sposobów adaptacji nazw zapożyczonych autorstwa m.in. M. Malec, T. Milewskiego, A. Cieślíkowej, M. Karpluk, Z. Kalety, J. Safarewicza czy S. Urbańczyka.

Imiennictwo późniejszych okresów długo znajdowało się na marginesie zainteresowań badawczych autorów opracowań antroponomastycznych, którzy w odniesieniu do wieków XVI–XX zdecydowanie częściej zajmowali się nazwiskami, które na przestrzeni tego okresu się kształtowały. Przez dłuższy czas, jak pisze Z. Abramowicz (1993: 6), fundamentalną pracą dotyczącą współczesnego polskiego imiennictwa była wydana w 1938 r. *Księga imion w Polsce używanych* J. S. Bystronia, zawierająca bogaty zbiór imion z różnych regionów, wyekscerpowanych z wybranych materiałów źródłowych z lat 1390–1921 oraz pierwsze zestawienia dotyczące frekwencji poszczególnych jednostek antroponimicznych i jej zmienności w czasie. Imiennictwem zajmowano się głównie w ramach (lub na marginesie) badań etnograficznych czy dialektologicznych. Do najistotniejszych tekstów tego nurtu M. Malec zalicza powojenne artykuły A. Zaręby (1957, 1959), w których zarysował on program badania nad polskimi imionami ludowymi, a także regionalne opracowania z obszaru Orawy T. Gołębiowskiej (1971) i okolic Niska M. Karasia (1969) oraz szczegółowe artykuły poświęcone zagadnieniom słowotwórstwa W. Dobrzyńskiego (1974) i wieloimienności A. Spyta (1973). Warto wspomnieć też o dwóch dużo starszych artykułach o charakterze przyczynkowym na temat imiennictwa ludowego w ziemi łęczyckiej M. Gralewskiego (1899) oraz w Gorcach S. Flizaka (1937). Niewielkie zainteresowanie problematyką imion nowszych warstw chronologicznych, a zwłaszcza imiennictwa współczesnego wynikało z przekonania, że bezpośrednim zwierciadłem kultury są tylko młode, pierwotne systemy imiennicze (Milewski 1963: 40) oraz z dominującego do niedawna w onomastyce paradygmatu strukturalistycznego. Zasadniczy przełom dokonał się w latach 70. XX w. wraz z dojściem do głosu paradygmatu socjolingwistycznego, skupionego na zależnościach między sferą nazewniczą a czynnikami pozajęzykowymi – kontekstem społecznym, kulturowym czy historycznym. Zwiastunem nowego podejścia badawczego do imiennictwa było ukazanie się artykułu S. Gajdy *Socjologia imion na przykładzie Opola i okolicy* (1973), w którym przy pomocy m.in. metod kwantytatywnych wykazał wewnętrzne zróżnicowanie (i jego zmienność w czasie) systemu imienniczego na badanym obszarze w latach 1949, 1959, 1969 warunkowane czynnikami takimi jak pochodzenie społeczne i geograficzne rodziców (ludność napływowa, ludność rodzima), płeć nosicieli imion czy względy estetyczne. W opublikowanym kilka lat później artykule poświęconym socjolingwistycznym aspektom imiennictwa H. Borek zwrócił uwagę na potrzebę badania zwłaszcza zaniedbywanego dotychczas imiennictwa współczesnego. Autor, zgadzając się z przytoczonym poglądem T. Milewskiego, zwrócił uwagę na konieczność stosowania odmiennego aparatu metodologicznego w badaniach materiału imienniczego różnych epok:

„Inne zadania muszą mieć badania diachroniczne, gdyż powinny one w pierwszym rzędzie wyjaśnić genezę poszczególnych systemów imiennicznych na tle ówczesnych stosunków społeczno-kulturowych, a także prześledzić ich późniejsze nawarstwienia i przekształcenia aż do typowego skostnienia tych systemów włącznie. (...) Na tym etapie rozwojowym odbicia zjawisk pozajęzykowych należy szukać przede wszystkim poprzez interpretację etymologiczno-znaczeniową imion. Natomiast z chwilą zaniku pierwotnej motywacji imion, tj. ich roli społecznej zgodnej z dawną treścią semantyczną, można już tylko śledzić ich żywotność i częstość występowania zależnie od zmiennych czynników zewnętrznych, nie motywowanych już przez elementy sprawcze, które powołały dany system imienniczny do życia, np. zależność od mody, tradycji rodzinnej, wzorów literackich itp.” (Borek 1978: 166)

Według autora na odmienny sposób traktowania imiennictwa nowszych okresów powinna wpływać także o wiele większa dostępność pełnego materiału źródłowego dla danej miejscowości czy szerszego obszaru niż w przypadku średniowiecza. Pozwala to na kompleksową analizę zasobu imion całej populacji w przekrojach pokoleniowych, klasowo-zawodowych, etnicznych czy wyznaniowych (Borek 1978: 167).

Postulaty te zaowocowały powstaniem coraz liczniejszych opracowań regionalnych, w których, poza dokumentacją zasobu imienniczego na określonym obszarze, podejmuje się w różnym zakresie zagadnienia dotyczące: częstości występowania imion, zmienności ich popularności w czasie, wariantowości form, motywów nadania czy związków z uwarunkowaniami socjalnymi. Warto nadmienić, że wśród autorów tego typu prac znajdziemy nie tylko językoznawców-onomastów, ale także przedstawicieli nauk historycznych. Imiennictwu z dłuższych lub krótszych okresów z przedziału XVI–XIX w. w znacznej mierze poświęcone są opracowania antroponimii zarówno pojedynczych miejscowości, m.in. Puław (Walczak 1980), Łasku (Kamińska 1986), Debrzna (Surma 1986), podkrakowskiej Rudawy (Klisiewicz 1989, 1994), Nowej Brzeźnicy k. Częstochowy (Tomecka 1994), śląskiego Olesna (Magda-Czekaj 2000), Białegostoku (Jaroszuk 1983), Zarzecza k. Niska (Maślach 2007), Kobióra k. Tychów (Ryguła 2011), podlubelskich Fajslawic (Adamiak 2013), Krakowa (Rudnicka-Fira 2013), parafii Stoczek na pograniczu mazowiecko-podlaskim (Kresa 2013), Poznania (Banderowicz 2015) czy parafii Pruchnik k. Jarosławia (Romanik, Kiper 2016), jak i szerszych obszarów: Wielkopolski (Skulina 1988), kresów południowo-wschodnich (Wolnicz-Pawłowska, Szulowska 1998) czy Mazowsza (Szulowska 2004)¹³, przy czym w pracy M. Walczaka (częściowo też u M. Magdy-Czekaj) uwzględniany jest również materiał XX-wieczny. Trzeba zresztą zauważyć, że w coraz mniejszym stopniu kładziony jest nacisk na badanie antroponimii podług cezur czasowych

¹³ Przedsięwzięciem mającym na celu syntetyczne ujęcie antroponimii doby średniopolskiej, uwzględniającym w ograniczonym stopniu również imiona, jest wydany w latach 2007–2016 słownik *Antroponimia Polski od XVI do końca XVIII wieku* pod red. A. Cieślakowej (od II tomu przy współpracy K. Skowronek), w którym duży nacisk kładzie się na frekwencję, chronologię i geografę poświadczeń antroponimów, oraz status społeczny ich nosicieli.

stosowanych w periodyzacji rozwoju polszczyzny. Ustalenie punktów czy przedziałów czasowych, w których bada się imiennictwo, warunkowane jest bądź dostępnością materiału badawczego, bądź też ma związek z różnymi czynnikami ekstralingwistycznymi – historycznymi, politycznymi czy administracyjnymi. Dobrym tego przykładem jest obszerna praca Z. Abramowicz (1993) poświęcona imionom białostoczan w latach 1885–1985, które autorka analizuje w trzech przekrojach chronologicznych – przynależności Białegostoku do Cesarstwa Rosyjskiego, okresu międzywojennego oraz PRL. Regionalne badania imiennictwa w węższych lub szerszych wycinkach czasu od przełomu XIX i XX po czasy współczesne, poza wymienionymi już pracami S. Gajdy, M. Walczaka i Z. Abramowicz, były prowadzone na materiale pochodzącym m.in. z Krakowa (Bubak 1978, 1983a, Parzniewska 1996), Bielska-Białej (Bubak, Mrózek 1983), Działoszyna i Łodzi (Umińska 1983, Umińska-Tytoń 1987, 1988), Bydgoszczy (Breza 1987), Łowicza (Plichta 1998), Częstochowy (Bizior-Ociepa, Jaros 2002), Łobżenicy (Czaplicka 2008, Czaplicka-Jedlikowska 2008), wybranych miast Górnego Śląska i Zagłębia Dąbrowskiego (Rudnicka-Fira 2011), Torunia (Zielińska 2012), Korczówki k. Białej Podlaskiej (Kuchta 2013), Drobin (Włodarski 2015), Gniezna (Iglewska 2015), Gdeszyna k. Zamościa (Józwiak 2016), Bielska Podlaskiego (Mordań 2017) czy Rzeszowa (Rejman 2017).

Na wzrost zainteresowania tematyką współczesnego imiennictwa duży wpływ miało także wydanie w 1995 r. przez K. Rymuta wspomnianego już w poprzednim rozdziale *Słownika imion współcześnie w Polsce używanych*. Ten obszerny wykaz umożliwił przeprowadzenie dotąd właściwie nieobecnych badań nad imiennictwem w skali ogólnopolskiej. Na jego podstawie powstały istotne prace dotyczące najnowszych przemian w polskim imiennictwie (do roku 1994 r.) w zakresie innowacji leksykalnych i gramatycznych (Bubak 1998, Malec 1998, 2007) oraz frekwencji (Nowik 1998). Materiał dla okresu 1995–2010, również pochodzący z PESEL, stanowił bazę dla serii przeglądowych artykułów dotyczących oficjalności i wariantywności imion (Cieślíkowa 2013), imion wielkomiejskich (Skowronek 2013), frekwencji i geografii imion rzadkich (Czopek-Kopciuch 2013) i częstych (Swoboda 2013). Mimo coraz większej dostępności wielkoskalowego materiału badawczego obejmującego imiennictwo całej Polski (zob. rozdział I.1), wciąż przeważają jednak opracowania wycinkowe, skupione jedynie na antroponimach jednej czy kilku miejscowości.

Słownik K. Rymuta stanowi też swego rodzaju chronologiczną kontynuację wspomnianych już pomnikowych prac leksykograficznych z zakresu antroponimii, takich jak *Słownik staropolskich nazw osobowych* czy *Antroponimia Polski od XVI do końca XVIII w.*,

choć znacząco się od nich różni pod względem sposobu opracowania materiału, zawierając jedynie dane o frekwencji i geografii poszczególnych jednostek antropimicznych bez jakichkolwiek objaśnień samych imion. W tym miejscu należy wspomnieć o antropimikonach o charakterze popularno-naukowym, powstałych w ostatnich dziesięcioleciach, które uzupełniają tę lukę i znacząco wzbogacają naszą wiedzę o współczesnym polskim imiennictwie. Mowa tu przede wszystkim o pracy J. Bubaka *Księga naszych imion* (1993), w której autor zestawia informacje o ponad 1320 imionach używanych dawniej lub współcześnie w Polsce, obejmujące ich pochodzenie językowe, pierwotne znaczenie, najstarsze zaświadczenia źródłowe, opisowe dane o popularności (popularne, rzadkie itd.), przykłady znanych nosicieli, nazwiska i toponimy utworzone od danego imienia, obcojęzyczne odpowiedniki oraz dzień imienin. Takie samo ujęcie prezentuje *Słownik imion* J. Grzeni (2002), zawierający charakterystyki ok. 1050 imion, użytych do nominacji w XX w. co najmniej 200 razy. Nie sposób pominąć też sześciotomowego opracowania H. Frosa i F. Sowy *Księga imion i świętych* (1997–2007), swoistego przewodnika onomastyczno-hagiograficznego po imionach chrześcijańskich w Polsce¹⁴.

Poza opracowaniami leksykograficznymi czy materiałowymi, które dokumentują i analizują współczesne imiennictwo w konkretnym zbiorze, w ostatnich dziesięcioleciach w obrębie polskiej onomastyki powstawały też teksty o charakterze bardziej ogólnym, dotyczące zagadnień teoretyczno-metodologicznych czy poprawnościowych. A. Cieślakowa (1988) skupiła się na społeczno-kulturowym aspekcie nadawania imion w Polsce. Poza wspomnianym już H. Borkiem (1978) zastosowaniu metod socjolingwistycznych w badaniach imiennictwa uwagę poświęcili W. Lubaś (1978), który pisał o użyciu imion w różnych społecznych typach kontaktu językowego, oraz J. Bubak (1987), który zaprezentował główne nurty w socjolingwistycznych badaniach antropimicznych w ówczesnym ZSRR. Temu ostatniemu autorowi zawdzięczamy też szereg prac na temat prawno-administracyjnych i poprawnościowych aspektów nadawania imion w Polsce (Bubak 1983a, 1983b, 1988).

Na zakończenie tego krótkiego przeglądu warto wspomnieć także, że polskie imiennictwo jako przedmiot badawczy zajmuje też przedstawiciele innych dziedzin niż językoznawstwo czy historia. Wśród przedstawicieli psychologii w pierwszej kolejności należy wymienić K. Doroszewicz, która jest autorką przeglądowej pracy przedstawiającej najistotniejsze

¹⁴ Praca ta stanowi rozwinięcie wcześniejszego dzieła obu autorów *Twoje imię. Przewodnik onomastyczno-hagiograficzny* (1976).

badania psychologiczne na temat imiennictwa prowadzone na świecie (Doroszewicz 2013). Jest ona też autorką – wspólnie z E. Stanisławiak – badań dotyczących stereotypów oraz psychologicznych wizerunków imion używanych w Polsce (Doroszewicz, Stanisławiak 1999), a także artykułów, w których pokazuje imiona jako narzędzie wpływu społecznego (Doroszewicz 2016) oraz opisuje – wspólnie z M. Krzemińską – sposoby wnioskowania o osobowości na podstawie formy imienia (Doroszewicz, Krzemińska 2007).

Imiona były też od dawna przedmiotem zainteresowania przedstawicieli nauk prawnych. Do najistotniejszych pozycji dotyczących prawnych aspektów imiennictwa należą dwa obszernie opracowania poświęcone imionom i nazwiskom oraz ich ochronie w polskim prawie autorstwa P. Ruczkowskiego (2010) oraz G. Tylca (2013). Trzeba jednak wiedzieć, że już dużo wcześniej (także wcześniej niż opracowania językoznawcze) powstawały prace mające na celu kompleksowe ujęcie prawnego funkcjonowania imion – najważniejsze wśród nich to: *Prawo imion* M. Allerhanda (1899) oraz *O imieniu i nazwisku. Studium prawne* J. Ostroróg-Sadowskiego (1902), w których zarysowano także historię kształtowania się polskiego systemu imienniczego. Z obszerniejszych studiów prawniczych warto wymienić także te poświęcone prawnym aspektom nadawania oraz zmiany imienia autorstwa A. Czajkowskiej (2007) i W. M. Hrynickiego (2010).

Symptomatyczne jest, że wspomniane opracowania i zagadnienia są stosunkowo mało znane w literaturze onomastycznej. I odwrotnie – z pobieżnej lektury wykazu prac wykorzystywanych przez wspomnianych prawników czy psychologów wynika, że w bardzo ograniczonym stopniu korzystają oni z dorobku polskiej onomastyki.

II.1.2. Socjolingwistyczne aspekty badań nad współczesnym imiennictwem

Spektrum zagadnień poruszanych w pracach nad imiennictwem uwzględniających jego ogólnie pojęte społeczne zróżnicowanie jest bardzo szerokie. Jego zakres w poszczególnych publikacjach jest różny – mamy do czynienia zarówno z opracowaniami, które kompleksowo analizują funkcjonowanie całych systemów imienniczych na rozległym tle czasowym, jak i bardziej lub mniej szczegółowymi studiami skupionymi na konkretnym problemie w wąsko zakreślonym wycinku czasowym. W niniejszej części pracy chciałbym poświęcić uwagę przedstawieniu głównych tendencji badawczych, jakie się uwidaczniają zarówno w polskich, jak i zagranicznych pracach nad imiennictwem najnowszych okresów.

Choć ujęcie socjolingwistyczne przejawia się często w badaniach antroponomastycznych, to tylko niektórzy badacze deklarują wprost stosowanie metod z tego zakresu w tytule pracy (np. Gajda 1973, Abramowicz 1993, Rudnicka-Fira 2011) lub w jej części poświęconej zagadnieniom metodologicznym (np. Umińska 1983: 133–134, Surma 1986: 182, Kresa 2013: 16). Na ogół zaś stosuje się je po prostu intuicyjnie, co wynika zapewne z tradycyjnego ukierunkowania onomastyki, które W. Lubaś określił mianem nurtu socjo-kulturowo-cywilizacyjnego (Lubaś 2013: 70).

Z polskich prac o charakterze metodologicznym (m.in. Borek 1978, Bubak 1987, Lubaś 2013), poświęconych stosowaniu metod socjolingwistycznych w onomastyce, wyłaniają się dwie różne perspektywy badawcze w odniesieniu do społecznego funkcjonowania imion, wynikające z odmiennego postrzegania celu i przedmiotu badania, ale także, a może przede wszystkim – z jego zakresu. Pierwsza z nich to perspektywa mikrosocjolingwistyczna, w ramach której akcent kładzie się na opis konkretnych użyci nazwy w akcie mowy z uwzględnieniem sytuacji, statusów i ról uczestników komunikacji (interakcji). Druga – makrosocjolingwistyczna – koncentruje się na opisach stratyfikacyjnych odmian języka przy użyciu pojęć odnoszących się do struktur ludnościowych, klasowych czy zawodowych (Bartol-Jarosińska 1991: 58), co jest też przedmiotem zainteresowania także w niniejszej pracy.

Perspektywa mikrosocjolingwistyczna reprezentowana jest w podejściu W. Lubasia (2013), który szczególny nacisk kładzie na komunikacyjny i pragmatyczny aspekt funkcjonowania nazw własnych. Przyjmując że podstawową kategorią socjolingwistyczną jest pojęcie wariantu i jego dystrybucja, zaś obiektem zainteresowania onomastyki są znaki językowe o jednostkowej denotacji, za przedmiot badania wariacji onomastycznej uważa przede wszystkim zróżnicowanie tekstowego użycia onimów w rozmaitych typach kontaktu

językowego. Każdy wyraz onomastyczny poza znaczeniem referencjalnym posiada bowiem także znaczeniowe składniki kontaktowe, odnoszące się do nadawcy lub odbiorcy komunikatu i jego stosunku do referenta. W tym ujęciu metody socjolingwistyczne służą do badania wariantów nazewniczych na poziomie uzusu, aktu mowy i towarzyszących mu okoliczności socjo-psychicznych. Przykładem tak pojmowanej wariacji onomastycznej może być użycie wyrazów antroponimicznych *Jan, Janek, Jasiulek, Jan Stanisławski, Stanisławski, Stan*, które są wariantami kontaktowymi oznaczenia tego samego referenta – stanowią dla siebie opozycje zależne od zmiennej sytuacji komunikacyjnej oraz różnych typów kontaktów kształtowanych m.in. przez stosunek jaki zachodzi między nadawcą a odbiorcą komunikatu (Lubaś 2013: 69–75).

H. Borek (1978), poza zarysowaną powyżej dostrzega też inną perspektywę badawczą, która plasuje się w obrębie makrosocjolingwistyki. Punktem wyjścia w analizie socjolingwistycznej jest tu funkcjonowanie całych systemów imienniczych w różnych grupach społecznych czy etnicznych oraz ich wzajemne zróżnicowanie, które warunkowane jest różnymi czynnikami pozajęzykowymi. Systemy te to nic innego jak zasoby imion (repertuary imiennicze) użyte w danej wspólnocie w akcie nominacji nowo narodzonego dziecka. Oczywiście nadanie imienia również stanowi akt komunikacji z towarzyszącym mu określonym kontekstem, na który składają się cechy zarówno nadawcy – rodziców, jak i odbiorcy – dziecka (ale też całej społeczności)¹⁵, mamy tu jednak do czynienia ze zredukowaniem sytuacji komunikacyjnej do jednego tylko przypadku – oficjalnego nadania imienia. Indywidualne decyzje rodziców w tym zakresie „wytyczają (...) w swojej masowości aktualny kształt systemów imienniczych i warunkują dalsze ich przekształcenia”. Badanie zarówno genezy, zmian i rozpowszechnienia poszczególnych systemów imienniczych, jak i śledzenie procesów decydujących aktualnie o określonym wyborze imion w różnych wspólnotach socjalnych i etnicznych jest tożsame z głównym zadaniem socjolingwistyki, jakim jest ukazanie związków między strukturą językową a strukturą społeczną (Borek 1978: 163).

Rozróżnienie tych dwóch perspektyw widoczne jest także w książce M. Knappovej (1989), która pisze o społecznych aspektach funkcjonowania imion w różnych sferach komunikacyjnych oraz w różnych grupach socjalnych. Na poziomie komunikacyjnym możemy mówić o dwóch stopniach, momentach funkcjonowania imion: a) w akcie nominacji,

¹⁵ A. Cieślukowa (1988: 163) zwraca uwagę na fakt, że nadanie imienia, w akcie mowy łączące się z użyciem czasownikami *nazywam (cię), chrzczę (cię)*, w wyniku aktu illokucji nie tylko skutkuje używaniem przez dziecko danego imienia, ale powoduje też, że imię to zostaje wprowadzone do świadomości społecznej.

gdzie badane są indywidualne motywy wyboru imienia; b) używania imienia, kiedy przedmiotem badania jest jego aktualna realizacja w mowie zachodząca w różnych typach kontaktu językowego (pośredniość, bezpośredniość komunikacji, jej oficjalność, półoficjalność, prywatność, relacje między uczestnikami itd.). Natomiast na poziomie społeczeństwa badaniu podlega to, w jaki sposób wybory imienia (których globalną sumę możemy postrzegać jako kolektywne zachowanie tworzące (pod)system imienniczy), odzwierciedlają stratyfikację terytorialną, socjalną, etniczną, wiekową itp. społeczeństwa. Przykładowe pytania stawiane w takim ujęciu dotyczą tego, czy istnieją różnice w nadawaniu imion i jego motywach pomiędzy poszczególnymi regionami kraju, między miastem a prowincją, w grupach o różnym statusie społecznym itd., a także czy zróżnicowanie to ulega zmianom w czasie (Knappová 1989: 50–59, 75–143).

J. Bubak (1987: 101) zwraca uwagę, że przyjęcie określonej perspektywy z jednej strony zależy od celów i przedmiotu badania, z drugiej zaś wiąże się z użyciem odmiennych technik gromadzenia materiału badawczego, które H. Borek (1978: 172) dzieli na subiektywno-ankietową (w skali mikro) i obiektywno-dokumentacyjną (w skali makro). Obie metody mają swoje wady i zalety, które zwięźle przedstawiła Z. Abramowicz (1993: 10, 26–28). Metoda ankietowa pozwala na ujawnienie całej złożoności subiektywnych motywów decydujących o wyborze danego imienia przez rodziców, a ponadto pozwala na ustalenie funkcjonowania imienia jako identyfikatora osoby w jego powszednim użyciu, w języku potocznym itd., czyli w różnych typach kontaktu językowego. Ma jednak ona zastosowanie tylko w badaniach nad imiennictwem współczesnym, w stosunku do osób urodzonych w chwili obecnej, dostępnej badaczowi – tylko osoby nadające imię (zwykle rodzice) mogą wyczerpująco odpowiedzieć na wszystkie pytania dotyczące motywacji wyboru imienia.¹⁶ Samo przeprowadzenie badań ankietowych na szerszą skalę nastęrcza trudności, toteż ogranicza się do niewielkiej próby w wąskim wycinku czasowym, sprowadzonym zwykle do jednego roku. Z tych oczywistych względów metoda ankietowa jest mało przydatna w badaniach imiennictwa w aspekcie diachronicznym, obejmujących szersze przedziały czasowe.

Do tego typu badań zdecydowanie bardziej odpowiednia jest metoda obiektywno-dokumentacyjna (czy też obiektywno-statystyczna), dla której materiał badawczy stanowią imiona i towarzyszące im dane przechowywane w określonych rejestrach (ich typy zostały przedstawione w rozdziale I.2.1). Dane te, nawet bardzo szczegółowe, w ograniczonym

¹⁶ Oczywiście badanie ankietowe można przeprowadzać również wśród starszych osób, jednak nie muszą one wiedzieć, jakimi motywami kierowali się ich rodzice przy wyborze imienia.

stopniu pozwalają badaczowi tylko na pośrednie wnioskowanie o subiektywnej motywacji wyboru imienia, i bardzo rzadko mówią cokolwiek o jego funkcjonowaniu w komunikacji. W dużo zaś większym stopniu umożliwiają ustalenie pełnego zasobu danego systemu imienniczego i przedstawienie jego zróżnicowania jakościowego i ilościowego w różnych przekrojach czasowych czy przestrzennych w kontekście czynników językowych, społecznych, kulturowych, historycznych czy politycznych, które to zróżnicowanie mogą determinować. Oczywiście tylko od badacza zależy w jakim zakresie dokona ekscerpcji materiału i w jaki sposób przedstawi wyniki jego analizy.

Prowadzenie badań nad współczesnym imiennictwem przy pomocy metody subiektywno-ankietowej postulował H. Borek, który w przytaczanym już artykule opublikował pierwsze wyniki analizy zgromadzonego materiału z Opola i Szczecina¹⁷ przy pomocy opracowanej przez siebie ankiety, która zawierała pytania o imię nadane dziecku, jego formy zdrobniałe używane w odniesieniu do niego, subiektywne czynniki decydujące o wyborze (tradycja rodzinna, współbrzmienie z nazwiskiem, brzmienie imienia i zdrobnień, data urodzenia – imię kalendarzowe, popularność lub rzadkość imienia, wpływ literatury lub mediów itp.), czy zostało nadane więcej niż jedno imię i jakie były tego powody, kto miał wpływ na decyzję o wyborze imienia, a także pytania dotyczące imion, miejsca zamieszkania, wiek rodziców, wykształcenie, zawód oraz pochodzenie rodziców. Autor skupił się przede wszystkim na subiektywnych motywach wyboru imienia – w obu miastach największy wpływ na wybór imienia miały czynniki eufoniczne – ładne brzmienie zdrobnienia oraz współbrzmienie z nazwiskiem, a także oryginalność imienia oraz tradycja rodzinna, zaś pozostałe czynniki miały znaczenie marginalne. Autor odnotował tu paradoks polegający na tym, że mimo deklarowania przez rodziców wyboru imienia rzadkiego, oryginalnego, w istocie wybierano imię popularne. Wiąże się to z faktem, iż trafny wybór rzadkiego imienia, które spodobało się w jakimś środowisku, wywołuje automatycznie mechanizm naśladownictwa, co sprawia, że imię to w krótkim czasie staje się powszechnym. Mniej uwagi autor poświęcił analizie repertuarów imiennicznych obu miast, ograniczając się do zestawienia list 15 najczęstszych imion męskich i żeńskich. Odnotował także ogólne zjawisko funkcjonowania w systemie imiennicznym podstawowego, tradycyjnego trzonu oraz marginalnego udziału w nim antroponimów oryginalnych, stanowiących innowacje leksykalne na polskim gruncie. W przypadku rodzimej ludności opolskiej H. Borek wspomina o zanikaniu imion o genezie

¹⁷ H. Borek referuje tu wyniki prac magisterskich K. Andrzejczak z lat 1974/1975 *Aspekt socjolingwistyczny funkcjonowania imion na przykładzie Opola i powiatu opolskiego* oraz E. Słomowicz z lat 1975/1976 *Motywacja imion noworodków z lat 1975/76 na przykładzie miasta Szczecina*.

germańskiej i analogicznym zjawisku dotyczącym imion wschodniosłowiańskich u ludności napływowej (1978: 168–175).

Zapewne z uwagi na trudności w gromadzeniu danych prace bazujące wyłącznie na badaniach ankietowych należą w Polsce jednak do rzadkości. Ich przykładem może być jeszcze artykuł E. Brezy (1987)¹⁸ poświęcony motywom wyboru imion dla dzieci urodzonych w Bydgoszczy w roku 1966 i 1976. Podstawę dla prezentowanych badań stanowiła przeprowadzona wśród rodziców 1521 dzieci¹⁹ ankietę opracowaną przez H. Borka, do której E. Breza dodał pytania o to, które imię (pierwsze czy drugie) używane jest w rodzinie, a które w środowisku szkolnym, a także od którego imienia pochodzi spieszczenie, którym określane jest dziecko w rodzinie. Podobnie jak w poprzednio omówionej pracy, najistotniejszym zagadnieniem były dla autora motywy wyboru imienia. Również i tu najważniejszym motywem towarzyszącym wyborowi imienia było brzmienie zdrobnienia, jednak następne motywy miały już inny układ niż w Opolu i Szczecinie. E. Breza ponadto porównał udział motywów w nominacji dziewcząt i chłopców (w ich hierarchii występują znaczne różnice), a także zmienność tych motywów w poszczególnych latach. Mniej uwagi poświęcił zagadnieniu wpływu czynników zewnętrznych, takich jak wiek czy wykształcenie rodziców, natomiast w ogóle nie uwzględnił kwestii frekwencji poszczególnych imion i jej zmienności w czasie. Poza samymi motywami autora interesuje też zagadnienie wieloimienności oraz zróżnicowanie formalne imion w różnych użyciach sytuacyjnych, zdecydowanie bogatsze w przypadku imion osób płci żeńskiej. Generalną tendencją odnotowaną przez E. Brezę jest fakt przywiązywania przez rodziców większej wagi do imion osób płci żeńskiej.

Badanie ankietowe zastosowano też w pracach J. Bubaka (1983), G. Surmy (1986) czy M. Czaplickiej (2008), jednak metoda ta nie posłużyła autorom do rejestracji imion oficjalnych, które pochodziły zwykle z ksiąg urzędów stanu cywilnego. J. Bubak zestawiał najczęściej nadawane imiona w Krakowie w latach 1975 i 1980 i zastanawiając się, czy imiennictwo krakowskie jest reprezentatywne dla innych regionów Polski w drugiej połowie lat 70., porównał je z listami dla Opoła i Szczecina, sporządzonymi przez H. Borka, oraz Cieszyna opracowaną przez R. Mrózka. Autor stwierdził istnienie dużej zbieżności (zwłaszcza w zakresie imion żeńskich) między poszczególnymi miastami. Z kolei ankietę przeprowadzoną wśród rodziców studentów J. Bubaka (urodzonych na przełomie lat 50. i 60.) dotyczyła motywów, którymi kierują się rodzice przy wyborze imienia. Dała ona nieco

¹⁸ Artykuł bazuje na wynikach pracy magisterskiej G. Jureko z 1981 r. *Socjolingwistyczne aspekty imiennictwa dzieci bydgoskich, urodzonych w latach 1966 i 1976*.

¹⁹ Tyle spośród 2200 rozesłanych ankiet zostało wypełnionych.

odmienne wyniki niż te zaprezentowane przez H. Borka. Według ankietowanych najistotniejsze były następujące motywy (w kolejności według liczby odpowiedzi): tradycja rodzinna, modność/popularność imienia/charakterystyczność imienia, imiona znanych postaci historycznych, współczesnych i literackich, imię przyniesione (kalendarzowe), brzmienie i forma imienia, kult świętego, intencje zawarte w imieniu (symbolika imienia), oryginalność/mała popularność, znaczenia imienia (etymologiczne). Ankieta ta dotyczyła jednak osób dużo starszych niż u H. Borka, co mogło zdaniem J. Bubaka mieć wpływ na zaobserwowane różnice.

G. Surma (1986) ustaliła zasób imion oficjalnych wszystkich mieszkańców gminy Debrzno na Pomorzu w latach 1977–1978 na bazie ksiąg meldunkowych i na tej podstawie obliczyła częstość użycia poszczególnych imion w dwóch grupach wiekowych od 0 do 30 lat oraz (powyżej 30 lat), rozumianych przez autorkę jako pokolenie młodsze i starsze. Pozwoliło jej to na porównanie repertuarów imienniczych w dwóch pokoleniach i wykazanie, które imiona cechuje ciągłość użycia, które odnoszą się tylko do pokolenia starszego (przykładowo od roku 1963 nie notowano już żadnych imion typowo niemieckich), a które w pokoleniu młodszym stanowią innowację. Natomiast w ramach eksploracji terenowej przeprowadzono badanie ankietowe, mające na celu zebranie imiennictwa potocznego – jak pisze autorka – faktycznie używanego przez ludność. Zestawienie go z imiennictwem metrykalnym ukazało znaczne zróżnicowanie form w obiegu oficjalnym i potocznym. W ankiecie pytano też o subiektywne motywy wyboru imienia²⁰, a także o miejsce urodzenia oraz o pochodzenie głowy rodu. Kwestie pochodzenia były istotne dla autorki, gdyż mieszkańcy Debrzna i okolic to, poza nieliczną grupą autochtonów, w głównej mierze ludność napływowa, przybyła po II wojnie światowej z różnych regionów kraju – jest to więc obszar językowo mieszany, co znalazło odbicie w obszernej analizie słowotwórczej form hipokorystycznych. Niestety, autorka nie ukazała w żaden sposób zależności między pochodzeniem regionalnym a wyborem imienia.

Podobny charakter ma artykuł M. Czaplickiej *Współczesne tendencje w nadawaniu imion dzieciom w Łobżenicy* (2008). Jego tytuł jest nieco mylący, bo autorka porównuje w nim zasoby imion odnotowane w księgach urodzeń w latach 1929 i 1990, uznając imiennictwo tych okresów za reprezentatywne dla pokolenia dziadków i wnuków. Oddzielne badanie ankietowe, uwzględniające motywy nadania imienia oraz to, kto decydował o wyborze

²⁰ Autorka nie zaznaczyła jednak, których osób dotyczyło pytanie o motywację – czy wszystkich, czy tylko rodziców dzieci urodzonych w Debrznie w czasie prowadzenia badań terenowych.

imienia, przeprowadziła jedynie wśród 100 respondentów – rodziców dzieci uczęszczających w latach 90. do jednego z łobzenickich przedszkoli.

Jak widać, badania ankietowe służą przede wszystkim ukazaniu subiektywnej motywacji imienia oraz różnic w jego funkcjonowaniu w obiegu oficjalnym i nieoficjalnym. Widzimy jednak, że w części prac w celu rejestracji pełnego zasobu imion danej społeczności i tak wykorzystuje się oficjalne rejestry, natomiast badanie ankietowe wykonywane jest niezależnie, niejako towarzysząco, zaś dane uzyskane tą drogą nie są kompatybilne z danymi metrykalnymi. Jak pisze Z. Abramowicz (1993: 10), wspólną cechą wielu powyższych prac jest zbyt jednostronne badanie materiału antroponimicznego (skupione tylko na jednym problemie) oraz ich ograniczony zakres czasowy, co wiąże się z pewnym ubóstwem badanych zasobów imion. Brak szerszych podstaw badawczych praktycznie uniemożliwia uchwycenie wielu zjawisk zachodzących w systemie imienniczym i określenia szerszego spektrum czynników warunkujących jego rozwój.

Oczywiście możemy odnieść to także do wielu prac o charakterze obiektywno-dokumentacyjnym, które choć czasem przedstawiają pełny zasób imiennictwa wyekscerpowany z różnych oficjalnych rejestrów, to często skupiają się na przypadkowo wybranym wąskim wycinku czasowym, a w analizie ograniczają się do pojedynczych zagadnień (por. np. Bubak 1978, Parzniewska 1996, Iglewska 2015). Teoretycznie możliwe jest zestawienie i porównanie wyników takich cząstkowych badań dotyczących różnych pojedynczych lat, obszarów geograficznych czy grup społecznych (co zresztą jest w niektórych pracach czynione), jednak z uwagi na wspomniany już brak skoordynowania badań oraz ich niespójność metodologiczną i tematyczną w praktyce jest często to bardzo utrudnione.

Przykładem pracy wykorzystującej metody obiektywno-dokumentacyjne, w której kompleksowo ukazano właściwości systemu imienniczego w kontekście zróżnicowania społecznego oraz ich przeobrażenia w czasie, jest przywoływany już artykuł S. Gajdy (1973). W trzech następujących po sobie punktach czasowych na pełnym materiale z lat 1949, 1959, 1969, pozyskanym z ksiąg urodzeń urzędów stanu cywilnego, bada on imiona nadawane w Opolu i jego okolicach. Takie przedziały czasowe nie zostały przez autora wybrane przypadkowo – pierwsza data odpowiada zakończeniu wzmożonego ruchu migracyjnego na Śląsku Opolskim, ostatnia zaś wiąże się z wejściem w wiek małżeński pierwszego pokolenia urodzonego i wychowanego na ziemiach zachodnich. Autor nie ogranicza się tylko do czysto statystycznej rejestracji opolskiego zasobu imion w tych latach, ale analizuje go pod kątem

bogactwa ilościowego (zob. rozdział II.1.3), repertuaru imion, wieloimienności, motywu wyboru imienia (dziedziczenie, zgodność z kalendarzem), dynamiki zmian w najczęstszym imiennictwie w obrębie różnych grup społecznych wyznaczanych takimi kategoriami socjologicznymi jak pochodzenie regionalne (ludność napływowa i rodzima), miejsce zamieszkania (Opole, powiaty namysłowski i oleski, opolski), status społeczny (chłopi, robotnicy, inteligencja wiejska, inteligencja miejska). Taka pogłębiona analiza doprowadza badacza do wniosku o współistnieniu dwóch głównych podsystemów imienniczych – ludności napływowej i rodzimej, będących wytworem odrębnego rozwoju historycznego germanizowanego Śląska i pozostałej części terytorium Polski, które w coraz większym stopniu podlegają integracji. Wyniki analizy wskazały też na istotne, choć również ulegające zacieraniu, różnice między imiennictwem żeńskim i męskim w zakresie dynamiki zmian i bogactwa ilościowego oraz dziedziczenia imion, świadczące o bardziej konserwatywnym podejściu przy wyborze imion męskich.

Do badań S. Gajdy, wykorzystując wypracowany przez niego aparat metodologiczny, nawiązała E. Umińska (1983) w artykule na temat imion metrykalnych mieszkańców okolic Działoszyna w woj. łódzkim oraz okolicznych wsi. Autorka analizuje w nim imiennictwo odnotowane w tamtejszych księgach urodzeń w wybranych latach: 1900 (okres zaboru rosyjskiego), 1920 (początek okresu niepodległej Polski), 1936 (ostatnie dostępne dokumenty z okresu przedwojennego), 1943 (okres okupacji hitlerowskiej), 1960 (początek budowy cementowni Warta, która spowodowała znaczne zmiany w strukturze ludności), 1979, oddzielonych na ogół 20-letnimi przedziałami odpowiadającymi mniej więcej jednemu pokoleniu. Autorka stwierdziła, że w coraz większym stopniu na wybór imienia obok tradycji rodzinnych i religijnych wpływ mają czynniki związane ze zjawiskiem mody. Analiza tempa zmian w zasobach imienniczych odpowiednich podsystemów doprowadziły badaczkę do wniosku o zwiększającym się wpływie mody miejskiej na zwyczaj nazewniczy wsi i zanikaniu tamtejszej tradycji imienniczej. Podobnie jak S. Gajda, autorka zauważyła, że zasób imion męskich jest znacznie bardziej konserwatywny od żeńskich – więcej jest w jego obrębie imion rodzimych, jest mniej otwarty na nowe jednostki imiennicze, w mniejszym stopniu ulega modzie na imiona obcojęzyczne. Autorka odnotowała też bardzo gwałtowne zmiany, jakie dokonały się w imiennictwie ludności polskiej okolic Działoszyna w okresie II wojny światowej, będące efektem polityki językowej hitlerowskiego okupanta, polegającej na administracyjnie wymuszonym nadawaniu dwuczłonowych imion słowiańskich (nierzadko sztucznych) w celu odróżnienia i swego rodzaju napiętnowania ludności polskiej.

W podobny sposób analizie poddane zostało imiennictwo Łodzi (Umińska-Tytoń 1987) – tu przedmiotem analizy były antroponimy odnotowane w latach 1900, 1920, 1938, 1960, 1979. Autorka odnotowała pewne różnice w repertuarach imienniczych dwóch odmiennych pod względem stopnia urbanizacji ośrodków – małomiasteczkowego Działoszyna i wielkomięskiej Łodzi, jednak wiele obserwacji pokrywało się w obu przypadkach, np. znaczny wzrost tempa zmian w obrębie najczęstszych imion czy bardziej konserwatywny stosunek rodziców do imion męskich. W imiennictwie łódzkim zaobserwowano ponadto zjawisko odnotowane już wcześniej przez H. Borka – mianowicie ścierania się dwóch przeciwstawnych tendencji – dążenia do nadawania imion oryginalnych z jednej strony, a z drugiej – wywołanego wpływem mody naśladownictwa i seryjnego nadawania tych samych imion. Zaskoczeniem okazał się znikomy wpływ obcych systemów antroponimicznych na polskie imiennictwo w warunkach wielokulturowości w latach 1900–1938, czyli okresie przemieszania ludności niemieckiej, polskiej i żydowskiej. Temat wpływu wydarzeń II wojny światowej na łódzkie imiennictwo został przez badaczkę rozwinięty w osobnym, obszernym artykule poświęconym imionom łódzian urodzonych podczas okupacji hitlerowskiej (Umińska-Tytoń 1988).

Celem badania E. Rudnickiej-Firy (2011) była z kolei odpowiedź na pytanie, czy Górny Śląsk i Zagłębie Dąbrowskie – sąsiadujące ze sobą obszary o silnym poczuciu tożsamości i odrębności regionalnej – utrzymują także odrębność w zakresie nadawania imion, przejawiającą się w jakościowym zróżnicowaniu ich repertuarów. Autorka poddała analizie materiał antroponimiczny dotyczący czterech pokoleń mieszkańców (z lat 1925, 1950, 1975, 2003), wyekscerpowany z ksiąg USC w czterech miastach Śląska (Katowicach, Chorzowie, Rudzie Śląskiej i Siemianowicach Śląskich) i trzech miastach Zagłębia (Sosnowcu, Czeladzi i Dąbrowie Górniczej), porównując 10 pierwszych miejsc na listach rangowych w każdym z nich w poszczególnych latach. W początkowej fazie na Śląsku występowała wyraźna przewaga imion o genezie niemieckiej, zaś w Zagłębiu występowały głównie imiona rodzime, dwuczłonowe. W kolejnym pokoleniu rozbieżności te są coraz mniej widoczne, co wiązało się głównie z odejściem od imion pochodzenia niemieckiego na Śląsku, zasadniczo jednak w obu regionach imiona były wybierane z dotychczasowych zasobów. W kolejnych badanych latach widoczna jest już nie tylko stopniowa unifikacja obu podsystemów, ale dochodzi też do wzmożonej rotacji w obrębie najczęstszych imion – w roku 2003 przeważają imiona nowe, nieznane lub bardzo rzadkie w poprzednich pokoleniach.

W przedstawionych powyżej pracach ewolucja lokalnych systemów imienniczych i ich przeobrażeń śledzona była na podstawie obserwacji w pojedynczych latach stanowiących swego rodzaju punkty pomiarowe. Z. Abramowicz (1993) w swojej obszernej monografii *Imiona chrzestne białostoczan w aspekcie socjolingwistycznym (lata 1885–1985)* stosuje nieco odmienną strategię. Jej celem jest bowiem przeprowadzenie kompleksowych badań na możliwe najbogatszym materiale imienniczym z Białegostoku (pozyskanym z ksiąg urodzeń i chrztów) na tle trzech okresów historyczno-politycznych, które pokrywa stuletni przedział zaznaczony w tytule: przynależności miasta do Cesarstwa Rosyjskiego, odzyskania polskiej państwowości (1919–1939) oraz okresu powojennego (1946–1985). Kategorią społeczną, która dla autorki stanowiła główną oś podziału zróżnicowania społecznego do 1939 r. była przynależność do jednej z przedwojennych grup wyznaniowych w Białymstoku – katolików, prawosławnych, protestantów i Żydów, które jednocześnie odpowiadały podziałom narodowościowym. Jak pisze Z. Abramowicz,

„nazewnictwo osobowe niezwykle silnie reagowało na wszelkie zmiany społeczne, historyczne i polityczne, w imieniu zawsze, niezależnie od woli nazywających, odzwierciedla się kultura i sytuacja socjalna społeczeństwa Białegostoku. Zestaw imion wykorzystanych na tym terenie w jakimkolwiek okresie nakreślał obraz społeczeństwa tego miasta” (Abramowicz 1993: 368).

Przykładowo autorka wykazała, że w pierwszym okresie w obliczu wzmożonej polityki rusyfikacyjnej poszczególne grupy narodowo-wyznaniowe starały się poprzez wybór imion z własnych systemów imienniczych podkreślić odrębność etniczną. Zmiana przynależności państwowej w okresie międzywojennym wiązała się z polonizacją urzędów, co znalazło odbicie także w imiennictwie: imiona katolików występują wyłącznie w wariantach polskich, zaś imiona mniejszości narodowych poddane były interferencji języka polskiego. W okresie PRL, wraz ze zmianą ustroju polityczno-społecznego oraz zmianą przekroju narodowościowego miasta, imiona obco brzmiące są już obligatoryjnie polszczone lub zastępowane wariantami polskimi. Obserwowana jest konsolidacja systemów imienniczych i ich stopniowa unifikacja. Wcześniejsze zróżnicowanie etniczno-konfesyjne odbija się nie tylko w dyferencjacji repertuarów imienniczych w poszczególnych grupach, ale widoczne jest też w kontekście takich zagadnień jak wieloimiennosc. W okresie rosyjskim wieloimiennosc praktykowana była przez ludność protestancką, wśród której obyczaj ten stopniowo zanika w latach międzywojennych, zaś staje się modny u katolików i Żydów, po wojnie obejmuje też prawosławnych, zaś w ostatnich latach badanego okresu stopniowo zanika wśród wszystkich mieszkańców Białegostoku. Również motywy wyboru imienia w okresie przedwojennym nie były jednakowe we wszystkich grupach ludności, z czasem jednak uległy uniwersalizacji.

Badania Z. Abramowicz objęły jednak nie tylko zróżnicowanie jakościowe czy motywacyjne różnych podsystemów imienniczych Białegostoku – autorka skupiła się też na ich strukturze ilościowej i nakreśliła pewien model mechanizmów warunkujących przemieszczanie się imion między poszczególnymi klasami częstości imiennictwa. Autorka wyodrębniła trzy główne klasy frekwencyjne, które mają swoje funkcje – imiona bardzo popularne, osiągające w pewnych okresach wysoką frekwencję (ponad 2% wszystkich użyć), średnio popularne (1% < 2%) oraz bardzo rzadkie. Dwie pierwsze klasy stanowią trzon współczesnego imiennictwa – przyczyną jego trwałości jest tradycja kulturowa i narodowa, a także pewne wartości eufoniczne. Najpopularniejsze imiona cechują się dużymi możliwościami derywacyjnymi oraz odpowiednią strukturą (składają się z 2–3 sylab, a w przypadku imion męskich mają spółgłoskowy nagłos i wygłos, najczęściej ze spółgłoską twardą). Imiona rzadkie określają z kolei kierunki rozwoju systemu imienniczego i ilustrują poszukiwania przez kolejne pokolenia nowych sposobów na jego wzbogacenie. W obrębie tego zasobu następuje systematyczna fluktuacja – do systemu wciąż wchodzi pewna grupa imion nowych, pewne z nich stopniowo przedostają się do klasy imion najczęstszych, następnie zaś przechodzą z powrotem do klasy nazw rzadkich. Droga ta ma może mieć różny przebieg – część imion stopniowo pnie się w górę i w taki sam sposób traci popularność, w przypadku innych zaś odbywa się to błyskawicznie. Jak pisze w podsumowaniu swojej pracy Z. Abramowicz, „imię wypełniając szereg socjalnych funkcji żyje i rozwija się według zasad języka, chociaż przyczyny stymulujące rozwój systemu imienniczego znajdują się poza sferą językową. Nawet forma imienia uzależniona jest od zjawisk pozajęzykowych” (Abramowicz 1993: 380–381).

Zagadnienia podejmowane przez polskich badaczy oraz obserwowane przez nich zjawiska, wynikające z wpływu rozmaitych czynników pozajęzykowych w polskim imiennictwie, są także przedmiotem refleksji w pracach zagranicznych badaczy dotyczących innych obszarów językowych i kulturowych. Szerokiego przeglądu opracowań, w których funkcjonowanie systemów imienniczych, ich zróżnicowanie i przeobrażenia rozpatruje się na tle wydarzeń historycznych, interakcji społecznych, procesów globalizacyjnych oraz w kontekście działania mechanizmów mody dokonała K. Doroszewicz (2013: 27–47). Na przykładzie wybranych prac postaram się zarysować zakres zagadnień, które są w nich poruszane.

Wpływem religii i ideologii na wybór imion zajmował się J. Gerhards (2005), porównując na przestrzeni lat 1894–1994 imiennictwo w dwóch miasteczkach – katolickim Gerolstein, znajdującym się na obszarze dawnych Niemiec Zachodnich, oraz Grimma we wschodniej

części Niemiec, w którym przeważali protestanci. Odrębne podejście w obu wspólnotach religijnych do kultu świętych zaznaczyło się w ich zachowaniach nazewniczych – w całym stuleciu imiona chrześcijańskie stanowiły w Gerolstein aż 70% wszystkich imion, podczas gdy w Grimmie było to jedynie 40%. Jednak rozwój ekonomiczny i postęp edukacyjny, które wiązały się rozluźnieniem związków z religią, a także inne procesy sekularyzacyjne, spowodowały, że w obu społecznościach odsetek ten, obserwowany w kolejnych latach, znacząco zmalał (z 50% do 32%). W okresie narodowego socjalizmu (i powiązanego z nim antyklerykalizmu) imiona chrześcijańskie znacząco zostały wyparte przez imiona germańskie (u protestantów, przywiązanych do idei narodowościowej, częściej nadawane były już od XIX w.). Wraz z zakończeniem wojny i kompromitacją ideologii nazistowskiej nastąpiła również ucieczka od kojarzonych z nią imion germańskich, wskutek czego ich udział w imiennictwie zmalał z 50% do zaledwie 5% w latach dziewięćdziesiątych. Od lat 50. niezależnie od ustroju politycznego panującego w obu częściach podzielonych Niemiec, obserwuje się gwałtowny wzrost imion obcego pochodzenia, co wiąże się z procesami globalizacyjnymi, wywołanymi przede wszystkim oddziaływaniem środków masowego przekazu – telewizji i muzyki pop, a także wzmożonym ruchem migracyjnym.

Przykładami prac, w których bada się zróżnicowanie imiennictwa w kontekście wpływu takich czynników jak pochodzenie etniczne/rasa, płeć oraz przynależność do danej warstwy społecznej, są artykuły H. Barry’ego III i A. Harper (2010) oraz G. Bloothoofa i D. Onlanda (2011). Bloothoof i Onland skupili się na pochodzeniu i językowej formie imion, które wybierane były w Holandii w latach 1982–2005 w poszczególnych grupach społecznych określanych przez takie wskaźniki socjoekonomiczne, jak dochód w rodzinie, poziom edukacji rodziców, a także styl życia, które bardzo często wiążą się też z pochodzeniem etnicznym. Rodzice o niższych dochodach oraz poziomie wykształcenia preferowali imiona angielskie, włoskie, hiszpańskie czy też „internacjonalizmy” imiennicze, natomiast preferencje osób zamożniejszych i lepiej edukowanych obejmowały imiona holenderskie, fryzyjskie, nordyckie, hebrajskie i francuskie. Zauważono również, że rodzice obyczajowo konserwatywni i przywiązani do wartości religijnych wybierają zwykle imiona tradycyjne, podczas gdy osoby o poglądach liberalnych, prowadzące bardziej nowoczesny styl życia są bardziej otwarte na nowe imiona, często charakteryzujące się krótką formą.

Z kolei Barry III i Harper analizowali bogactwo ilościowe imion nadawanych w latach 1990, 1995, 2000 w Pensylwanii w USA w obrębie czterech grup zdefiniowanych jako „białe kobiety”, „czarne kobiety”, „biali mężczyźni”, „czarni mężczyźni”. Okazało się, że

najbardziej różnorodne imiennictwo występowało wśród „czarnych kobiet”, zaś najuboższy ilościowo był zasób imion „białych mężczyzn”. Porównanie ilościowej zmienności repertuarów imienniczych w czasie wykazało jednak, że stopniowo zacierają się różnice pomiędzy imiennictwem w podziale na płeć czy rasę.

Kwestia zróżnicowania zachowań nazewniczych ze względu na płeć osoby otrzymującej imię, która pojawiła się w powyższym artykule, a także we wcześniej omówionych pracach polskich onomastów, była też przedmiotem zainteresowania D. Gerritzen (1998/99), która zbadała zmiany wzorców imienniczych dziewcząt i chłopców w Holandii w XX wieku na tle przemian kulturowych. Autorka porównała materiał imienniczy z lat: 1870–1947 i 1980–1995 pod kątem nadawania imion derywowanych od imienia pierwotnie używanego w odniesieniu do osób płci przeciwnej (ang. *transgendered names*), zawierających sufiks deminutywny (-je(n), -ke(n)), pochodzenia językowego nadawanych imion oraz motywacji towarzyszącej wyborowi imienia (tu częściowo opierając się na badaniach sondażowych). W pierwszym okresie aż ok. 50% dziewcząt otrzymywało imiona utworzone według modelu *imię męskie + feminizujący sufiks*, podczas gdy liczba chłopców noszących imię derywowane od imienia żeńskiego stanowiła zaledwie 0,5% całej męskiej populacji. Tak wielkie różnice między płciami obserwowane były także w przypadku nadawania imion z sufiksem deminutywnym, które dotyczyły 30% dziewcząt (taki był też udział deminutywnych imion w słowniku imion żeńskich) i tylko 1,1% chłopców (5% słownika imion męskich).

Zjawisko częstszego tworzenia imion żeńskich od męskich zwykle jest postrzegane jako przejaw patriarchalnego modelu kultury czy koncepcji „żebra Adama”, jednak autorka skłania się ku teorii o istnieniu znacznie silniejszych stereotypów płciowych na temat mężczyzn niż kobiet. Pod koniec XX wieku zróżnicowanie w tym zakresie zanikło, jednak odmienne postrzeganie ról społecznych kobiet i mężczyzn może nadal ujawniać się w innym aspekcie. Imiennictwo żeńskie w tym czasie uległo „feminizacji” (używane są imiona pierwotnie żeńskie) w głównej mierze w związku z częstszym obecnie nadawaniem dziewczętom imion zapożyczonych (zwłaszcza z języka angielskiego, w którym dużo mniej jest żeńskich derywatów od imion męskich), podczas gdy chłopcy otrzymują w głównej mierze imiona tradycyjnie holenderskie (zwykle proste i krótkie). Różnica dotyczy też motywacji wyboru imienia – dziedziczenie imienia po przodkach częściej występowało i występuje wśród chłopców, natomiast w przypadku imion żeńskich istotniejszą rolę pełnią obecnie względy estetyczne. Podobne zjawiska zaobserwowali wcześniej też inni badacze (por. np. Anderson 1985, Lieberson, Bell 1992) – imiennictwo żeńskie jest bardziej podatne na zmiany i modę

niż męskie, chłopcom nadawane są częściej imiona wyrażające ciągłość i tradycję, kojarzone z trwałymi wartościami.

Centralnym zagadnieniem w pracach nad imiennictwem amerykańskiego socjologa S. Lieberona jest z kolei kwestia mody imienniczej. Istotą mody, w imiennictwie rozumianej jako duża popularność jakiegoś zasobu antroponimów w określonym czasie, jest zmiana, która dotyczy szeroko pojętych aspektów estetycznych dotyczących jakiegoś przedmiotu lub pojęcia. Badanie tej zmienności na przykładzie imion ma dla wielu badaczy zaletę wynikającą z faktu, iż w odróżnieniu od wielu innych obszarów, gdzie obserwuje się działanie mody, wybór danego imienia nie jest podyktowany zasobami materialnymi i może stanowić jego obiektywny miernik.

W swojej obszernej rozprawie *A Matter of Taste. How Names, Fashions, and Culture Change* (2000) S. Lieberon, którą uzupełnia też późniejszy współautorski artykuł (Lieberon, Lynn 2003), dokonał analizy ilościowych i jakościowych przeobrażeń imiennictwa w ośmiu krajach Europy i Ameryki Północnej w kolejnych latach na przestrzeni co najmniej dwóch wieków (od wieku XX wstecz) i zaobserwował, że istotne zmiany w zakresie wymiany zasobów imienniczych oraz ich koncentracji zaczęły dokonywać się dopiero na przełomie XIX i XX w., a w kolejnych tempo i rozmiar tych zmian ulegały zwiększeniu. Wcześniej, zdaniem badacza, istniały określone reguły nadawania imion, które odtwarzały się w kolejnych pokoleniach, rozkłady częstości imion w jednym pokoleniu dzieci nie odbiegały w sposób znaczący od imion w pokoleniu rodziców.

S. Lieberon, zastanawiając się nad przyczynami, które sprawiły, że obowiązujący zwyczaj nazewniczy został zastąpiony indywidualnymi gustami rodziców, a więc zaczął podlegać mechanizmom mody, opisał dwie grupy czynników – zewnętrznych i wewnętrznych, które te mechanizmy warunkują. Czynniki zewnętrzne obejmują: dostęp do edukacji, procesy urbanizacyjne, zmiany w modelu rodziny oraz rozpowszechnienie się mass-mediów. Badania Lieberona pokazały, że zachodzi wysoka korelacja w czasie pomiędzy wskaźnikami takimi jak tempo rotacji czy zmniejszanie się koncentracji w imiennictwie (zob. rozdział II.3) a wskaźnikami dotyczącymi wymienionych czynników. Wewnętrzne mechanizmy mody, które sterują zmianami popularności imion, związane m.in. z ich symboliką oraz formą językową (brzmieniem, strukturą itp.), Lieberon określił obrazowo jako „efekt zapadkowy” (ang. *ratchet effect*). Imiona popularne niemal zawsze torują drogę do popularności innym, podobnie brzmiącym. Zmiany w imiennictwie zachodzą powoli, stopniowo i zawsze w powiązaniu z poprzedzającym stanem rzeczy – nowe imiona w pewnym sensie zazębiają

się z tymi, które były popularne dotychczas. Mechanizm ten jest uniwersalny i zachodzi także w innych obszarach – Lieberson porównuje go np. z modą w zakresie ubioru. Podobnie jak spódniczka mini nie pojawiła się od razu po długich spódnicach – stopniowo były one skracane, zawsze o tyle, na ile była to społecznie akceptowalna zmiana, tak samo nowe grupy imion nie mogły drastycznie odbiegać od imion dotychczas popularnych lub zakorzenionych w systemie antroponimicznym.

Powyższa obserwacja znalazła potwierdzenie w badaniach m.in. J. Gerhardsa (2005: 95–98) i P. Swobody (2013: 44–53), którzy wykazali, że zwiększenie lub zmniejszenie popularności określonego imienia często pociąga za sobą nieco przesunięte w czasie takie samo zachowanie imion reprezentujących podobne cechy brzmieniowe. Kompleksowo zagadnienie to opisała grupa amerykańskich badaczy (Berger i in. 2012). Na podstawie analizy trendów w popularności wszystkich imion nadawanych w USA w latach 1882–2006 zaczerpniętych z SSA stwierdzili oni, że im więcej wspólnych fonemów ma para imion, tym silniejsza jest korelacja ich popularności w czasie. Dowodzi to działania w kolektywnych zachowaniach nazewniczych znanego z psychologii efektu czystej ekspozycji, polegającego na tym, że im częściej się z kimś (lub czymś) stykamy, tym nasz stosunek do niego jest bardziej pozytywny. Ekspozycja (w przypadku imienia – jego popularność) zwiększa preferencje nie tylko eksponowanego bodźca, ale też bodźców podobnych, mających z nim cechy wspólne (Doroszewicz 2013: 31, przyp. 1).

Abstrahując od zagadnień teoretycznych podejmowanych we wspomnianych pracach S. Liebersona, należy zwrócić uwagę na bardzo ważną, także z punktu widzenia niniejszej pracy, kwestię sposobu analizy materiału imienniczego, jaką ten badacz jako jeden z pierwszych zastosował na szeroką skalę. W dużej części jego badania koncentrują się bowiem na analizie zmienności w czasie cech ilościowych poszczególnych systemów imiennicznych, takich jak stopień koncentracji (w ujęciu Liebersona jest to pokrycie przez pewną liczbę najczęstszych jednostek globalnej sumy użyć imion) czy wskaźnik tempa rotacji (liczba różnic między zasobami imion w kolejnych latach). Oczywiście sprawy te były przedmiotem zainteresowania w wielu innych wcześniejszych pracach, jednak dopiero u Liebersona mamy do czynienia z systematycznym i szczegółowym badaniem na pełnym materiale imiennicznym wielu krajów (lub ich regionów), obejmującym poszczególne lata z szerokiego przedziału czasowego. Pozwala nam to w skali globalnej, niejako „z oddali”, prześledzić, w jaki sposób i kiedy dokonywały się zmiany ilościowe w imiennictwie i uchwycić konkretne momenty zwrotne, a także określić, czy w takim samym stopniu

dotyczy to systemów imienniczych różnych obszarów. W ostatnich dwóch dziesięcioleciach, wraz z pojawieniem się możliwości dostępu do wielkoskalowych baz danych obejmujących imiennictwo całych krajów oraz z rozwojem narzędzi informatycznych pozwalających na ich analizę, zaczęły powstawać prace, których głównym celem jest badanie czasowego i przestrzennego zróżnicowania imiennictwa przez pryzmat pewnych cech kwantytatywnych, które je charakteryzują. Bardziej szczegółowo odniosę się do nich w kolejnym rozdziale (II.1.3), tu zaś ograniczę się do kilku przykładów ważnych dla mnie prac na temat imiennictwa w Stanach Zjednoczonych i zagadnień w nich poruszanych.

Celem artykułu J. M. Twenge, E. M. Abebe i W. K. Campbell (2010) było sprawdzenie, czy w zakresie zachowań imienniczych przejawia się tendencja do indywidualizacji i podkreślania unikalności, zauważana w wielu innych dziedzinach życia. W tym celu prześledzili oni jak w poszczególnych latach z przedziału 1880–2010 kształtował się odsetek dzieci (w podziale na płeć), którym nadano 1, 10, 25 i 50 najczęstszych imion w danym roku. Dane naniesione na wykresy pokazały, że na przestrzeni całego badanego okresu doszło do drastycznego spadku nadań najpopularniejszych imion, zaś współczynniki korelacji między odsetkiem najczęstszych imion i kolejnymi latami miały wysoką wartość ujemną. Z najbardziej gwałtownym spadkiem mieliśmy do czynienia w latach 1990–2010. Początkowa hipoteza zakładająca wpływ na odnotowane zjawisko wzmożonego napływu imigrantów zwłaszcza z krajów Ameryki Południowej w ostatnich latach (a wraz z nim zwiększenia się repertuaru imion) nie znalazła potwierdzenia w dalszej analizie porównawczej imiennictwa sześciu stanów o najmniejszym odsetku Latynosów wśród mieszkańców. Jej wyniki pokazały, że na obszarach tych tendencja jest taka sama jak w skali całego kraju. Okazało się natomiast, że stan Mississippi, w którym mieszka najwięcej osób czarnoskórych, na tle pozostałych stanów przez cały okres odznaczał się znacznie niższym odsetkiem nadań najczęstszych imion, co otwiera pole do dalszych eksploracji w zakresie zróżnicowania regionalnego w imiennictwie. Badania potwierdziły też opisywane przeze mnie wcześniej twierdzenia o bardziej zindywidualizowanym i otwartym na zmiany imiennictwie żeńskim – odsetek imion najczęstszych zawsze był w jego obrębie niższy niż w imiennictwie męskim. W artykule stanowiącym kontynuację opisanych badań (Twenge i in. 2016) potwierdzono pogłębianie się opisywanej tendencji indywidualistycznej w imiennictwie Amerykanów w kolejnych latach 2011–2016.

Regionalnego zróżnicowania w zakresie nadawania imion w USA dotyczyły przeprowadzone niezależnie od siebie badania dwóch zespołów – włoskiego (Barucca i in.

2015) i polskiego (Pomorski i in. 2015). W obu przypadkach autorzy postanowili zbadać, czy pewien ogólnie przyjęty tradycyjny podział geograficzny na Południe, Północny Wschód, Środkowy Zachód i Zachód (South, Northeast, Midwest, West), mający swoje podłoże historyczno-kulturowe, znajduje odzwierciedlenie także w imiennictwie. Obie grupy badaczy zastosowały w tym celu zbliżone, choć nieco odmienne procedury statystyczne, polegające na grupowaniu zbiorów imion z poszczególnych stanów na podstawie zbieżności w zakresie rozkładu częstości jednostek imienniczych w tych zbiorach. Miarą siły związku między parą stanów była siła korelacji między częstością kolejnych pozycji z list frekwencyjnych imion w tych stanach. Przeprowadzone tak obliczenia korelacji dla każdej pary stanów w poszczególnych latach dały bardzo ciekawe wyniki. Okazało się, że aż do przełomu lat 70. i 80. w zakresie imiennictwa utrzymywał się niemal sztywny układ geograficzny stanów pokrywający się z podziałem opisanym powyżej. Od tego momentu obserwowane jest załamanie się tego typu regionalizacji w imiennictwie i w kolejnych latach układ geograficzny oparty na coraz słabszej korelacji w zakresie rozkładu częstości imion ma bardziej chaotyczny i przypadkowy charakter, co może wiązać się z transformacją i modernizacją kulturową połączoną z unifikacją zachowań nazewniczych.

Przedstawiony w tym rozdziale wybiórczy przegląd polskich i zagranicznych prac na temat imiennictwa o nachyleniu socjolingwistycznym czy też socjoonomastycznym w żaden sposób nie wyczerpuje bogatej tematyki, jaka jest w tego typu badaniach poruszana. Miał on na celu przede wszystkim zarysowanie głównych tendencji badawczych oraz najważniejszych zagadnień, które się w nich pojawiają. Wiele z nich stanowiło bezpośrednią inspirację dla podjęcia badań przeprowadzonych w niniejszej pracy, dotyczących czasowej i przestrzennej wariacji systemu imienniczego w Polsce czy też jego zróżnicowania ilościowego imiennictwa ze względu na płeć nazywanej osoby.

II.1.3. Metody statystyczne w badaniach nad imiennictwem

W przeglądowym artykule dotyczącym polskich badań z zakresu statystyki językoznawczej M. Ruszkowski (2005: 54) odnotowuje tylko jedną pozycję z obszaru onomastyki – studium statystyczno-kognitywne na temat współczesnych nazwisk polskich autorstwa K. Skowronek (2001), w którym autorka zbadła m.in. działanie prawa Zipfa w tej sferze onimicznej. Choć książka ta jest rzeczywiście jedyną polską pracą onomastyczną, w której za główny przedmiot rozważań bierze się statystyczne właściwości zbiorów antroponimów, nie oznacza to jednak, że w polskiej onomastyce nie stosowano poza tym metod statystycznych. Należy jednak zaznaczyć, że odbywało się to w bardzo ograniczonym stopniu, nierzadko na marginesie. W niniejszym rozdziale skupię się na pracach tak polskich, jak zagranicznych, w których wykorzystano metody kwantytatywne w celu opisu danego zbioru imienniczego lub komparacji dwóch lub więcej systemów imiennicznych.

W niedawno opublikowanym artykule *Zastosowanie ilościowo-jakościowej metody statystyki lingwistycznej w badaniach siedemnastowiecznego zasobu chrześcijańskich imion osobowych północnej Rosji* B. Hryniewicz-Adamskich zwraca uwagę na powszechne zredukowanie rozumienia statystyki w badaniach antroponimicznych, polegające na tym, że „Autorzy większości publikacji poświęconych statystyce imion osobowych ograniczają się do prostego podliczenia ich ilości i wskazania najczęstszych spośród nich. Niekiedy tworzą listy frekwencyjne z uwzględnieniem procentowego udziału poszczególnych imion w ogólnej liczbie nominacji” (Hryniewicz-Adamskich 2018: 183).

Rzeczywiście w niemal wszystkich pracach, w których badaniu poddaje się jakiś szerszy zasób imiennictwa zawarty w danym rejestrze, nawet tych, których autorzy w tytule odwołują się do statystyki (por. np. Michałek 1995), stosowanie metod kwantytatywnych polega głównie na zestawieniu frekwencji (bezwzględnej lub względnej) poszczególnych jednostek antroponimicznych w zbiorach imion oraz określeniu udziału w nich różnie definiowanych grup imiennictwa i porównaniu pod tym kątem różnych zbiorów imion. Jak pisze B. Hryniewicz-Adamskich, walorem wykorzystania w badaniach antroponomastycznych ilościowo-jakościowej metody statystyki lingwistycznej V. D. Bondaletova (1983), która pojawia się w tytule jej pracy, jest to, że „pozwala ona w liczbach bezwzględnych określić «bliskość»/«oddalenie» porównywanych zasobów imiennicznych w perspektywie synchronicznej, a także diachronicznej” (Hryniewicz-Adamskich 2018: 184). Kolejne kroki przeprowadzonej przez autorkę analizy porównawczej zasobów imion z dwóch okresów poza stworzeniem dla nich list frekwencyjnych obejmują: wskazanie liczby imion wspólnych dla

obu zasobów oraz właściwych tylko jednemu z nich, określenie liczby ich nosicieli i ich procentowego udziału w obu zbiorach, określenie ile procent więcej nosicieli danego imienia wystąpiło w jednym z okresów (obliczenie różnicy), obliczenie średniej liczby użytkowników jednego imienia²¹, ustalenie procentowego udziału nominacji z użyciem imion częstych (powyżej średniej) oraz 5 i 10 najczęstszych imion i porównanie obu zasobów imienniczych pod tym kątem. Warto przypomnieć jednak, że w badaniach nad polskim imiennictwem tego typu analizy porównawcze, uwzględniające ilościowe zróżnicowanie zbiorów mierzone odpowiednimi statystykami opisowymi (nawet w szerszym niż zaproponowany tu zakresie), nie są niczym nowym i były stosowane dużo wcześniej.

Prekursorem takiego podejścia w polskiej antroponomastyce jest S. Gajda. W omawianym już artykule (Gajda 1973) autor omówił nie tylko różnice jakościowe pomiędzy zbiorami imion nadanych dzieciom urodzonym w latach 1949, 1959 i 1969, ale zwrócił także uwagę na pewne właściwości kwantytatywne, które je różnicują. Jednym z zagadnień, jakie poruszył badacz, była kwestia ilościowego bogactwa systemu imienniczego. Zauważył bowiem, że zbiory imion męskich i żeńskich, choć podobne pod względem sumy użyć imion (tj. osób, które otrzymały imię w danym roku), różnią się pod względem liczby różnych imion użytych do nominacji. Autor, zdając sobie sprawę, że nawet niewielkie różnice w sumie użyć mogą wpływać na liczbę imion, do zmierzenia bogactwa obu podsystemów (męskiego i żeńskiego) posłużył się miarą przeciętnego użycia jednego imienia, zakładając, że im niższa przeciętna, tym bogatszy system imienniczy²². Okazało się, że w poszczególnych latach w Opolu większe bogactwo cechowało zawsze imiennictwo żeńskie²³. Ponadto badacz porównał procentowy udział użyć najczęstszych 5, 10 i 15 antroponimów w globalnej sumie użyć wszystkich imion, co doprowadziło go do wniosku, że upodobania rodziców były bardziej skonkretyzowane w przypadku imion żeńskich (zwykle udział r najczęstszych imion był wśród nich wyższy), a także że stopień konkretyzacji z dekady na dekadę wykazywał tendencję spadkową. Autor dokonał też obliczeń udziału imiennictwa najrzadszego o frekwencji wynoszącej 1 i 2 (czyli hapaks- i dislegomenów), pozostawiając jednak ich wyniki bez komentarza. Przedmiotem rozważań S. Gajdy była również dynamika zmian w obrębie najczęstszego imiennictwa – zestawił on liczbę zgodności pomiędzy listami 10 najczęstszych imion w powiatach opolskim,

²¹ Autorka stosuje tu nieco mylący termin „współczynnik jednoimienności”, który może kojarzyć się ze zwyczajem nadawania jednego imienia (w opozycji do wieloimienności).

²² Czyli stosunku sumy użyć wszystkich imion do liczby unikatowych imion użytych do nominacji.

²³ Autor porównał te wyniki także z obliczeniami na podstawie danych z badań S. Folfasińskiego (1947–49) o frekwencji imion nadanych w Częstochowie i okazało się, że tendencja była tam odwrotna – przeciętna liczba użyć imion żeńskich była wyższa niż męskich.

namysłowskim i oleskim w następujących po sobie punktach czasowych, co ukazało tempo rotacji w obrębie najczęstszeo imiennictwa (zdaniem autora większa liczba zmian oznaczała większy wpływ mody na dobór imienia), a także stopień zbieżności między podsystemami imiennicznymi poszczególnych obszarów.

Metody wypracowane przez S. Gajdę zastosowała E. Umińska w badaniach imiennictwa mieszkańców okolic Działoszyna (Umińska 1983), porównując ze sobą zbiory imion (osobno męskich i żeńskich) odnotowanych w księgach metrykalnych parafii działoszyńskiej w latach 1900, 1920, 1936, 1960 i 1979 przy pomocy zaproponowanych przez niego wskaźników. Różnice w bogactwie między imiennictwem męskim i żeńskim były według badaczki nieznaczne, choć widoczna była pewna tendencja – w niewielkim stopniu zasób imion męskich był bogatszy w roku 1900, natomiast w kolejnych punktach czasowych sytuacja była odwrotna bądź też nie wystąpiły istotne różnice. Analiza udziału *r* najczęstszych imion w ogólnej sumie użyć przyniosła podobne spostrzeżenia, jak w przypadku badań S. Gajdy – dobór imion żeńskich był bardziej skonkretyzowany, jednak z czasem zarówno sam stopień konkretyzacji, jak i różnice w jego zakresie między imionami męskimi i żeńskimi ulegały systematycznemu zmniejszaniu. Porównanie list zgodności między poszczególnymi latami, pokazało z kolei, że tempo zmian w zasobie imion żeńskich (a także suma użyć nowych imion) jest większe niż w przypadku imion męskich, co również miało świadczyć o większej podatności na wpływ mody imiennictwa żeńskiego.

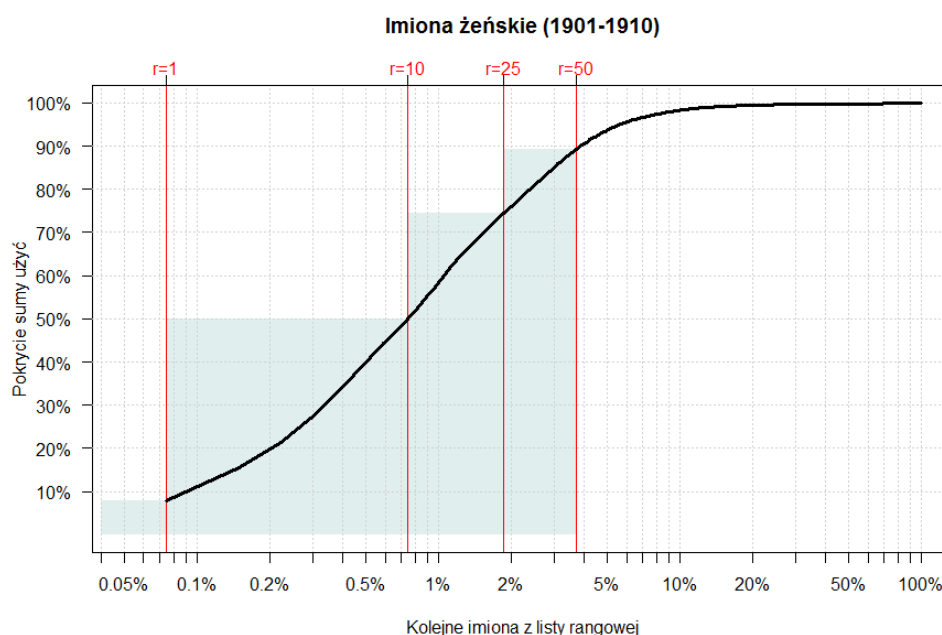
W wielu późniejszych tekstach dotyczących współczesnego imiennictwa różnych obszarów polscy badacze, choć zwykle odwołują się do wymienionych wcześniej prac S. Gajdy czy E. Umińskiej, to rzadko wykorzystują zastosowane przez nich metody statystyczne, ograniczając się co najwyżej do podania informacji o wielkości słownika imion w porównywanych zbiorach (por. np. Parzniewska 1996: 139, Iglewska 2015: 213–214). Nieco więcej uwagi strukturze liczbowej zbiorów imion poświęcili K. Nowik (1998), K. Skowronek (2001²⁴, 2013) i P. Swoboda (2013) w pracach poświęconych imiennictwu ogólnopolskiemu i wielkomiejskiemu. W tekstach tych przedstawiono obliczenia dotyczące pokrycia przez leksykę antroponimiczną różnych klas częstości globalnej sumy użyć wszystkich imion oraz słownika imion oraz do wielkości słownika imion. Ponadto K. Skowronek (2001: 86–87) w celu porównania zbiorów polskich nazwisk i imion posłużyła się wskaźnikami ilościowego bogactwa leksykalnego tekstu P. Guirauda i W. Kuraszkiewicza opartymi na stosunku między liczbą różnych wyrazów w tekście (tu wielkością słownika

²⁴ Na marginesie badania statystycznej struktury nazwisk polskich.

antroponimów) a długością tekstu (sumą użyć antroponimów), należących do grupy wskaźników *type-token-ratio (TTR)*²⁵.

Większość badaczy zajmujących się polskim imiennictwem, w choćby najmniejszym stopniu posługując się metodami kwantytatywnymi, odnotowywała pewną prawidłowość dotyczącą rozkładu częstości w zbiorach imienniczych. Polega ona tym, że stosunkowo niewielka liczba najczęściej używanych imion pokrywa ogromną część globalnej sumy użyć, a jednocześnie stanowi bardzo mały wycinek słownika imion. Odwrotna jest natomiast w przypadku antroponimów o najniższych frekwencjach – ich udział w całym słowniku jest bardzo duży, jednak stanowią one niewielką część ogólnej sumy użyć. Przykładowo na podstawie danych o frekwencji pochodzących ze *Słownika imion współcześnie w Polsce używanych* K. Rymuta (1995) K. Skowronek ustaliła, że liczba nadań 100 najczęstszych imion, które stanowią zaledwie 0,52% wszystkich 19 238 użytych imion, wyczerpuje 81%

Rys. 5. Pokrycie sumy użyć imion przez kolejne imiona z listy frekwencyjnej



globalnej sumy użyć, natomiast imiona nadane tylko jeden lub dwa razy stanowią aż 49,1% słownika imion i jednocześnie pokrywają tylko 0,051% sumy użyć wszystkich imion (Skowronek 2001: 198). Podobne zjawisko dobrze obrazuje wykres na rys. 5, na którym przedstawiono pokrycie sumy użyć imion żeńskich z arbitralnie wybranego okresu 1901–1910 z bazy PESEL (ok. 400 tys. okazów) przez kolejne imiona z listy frekwencyjnej

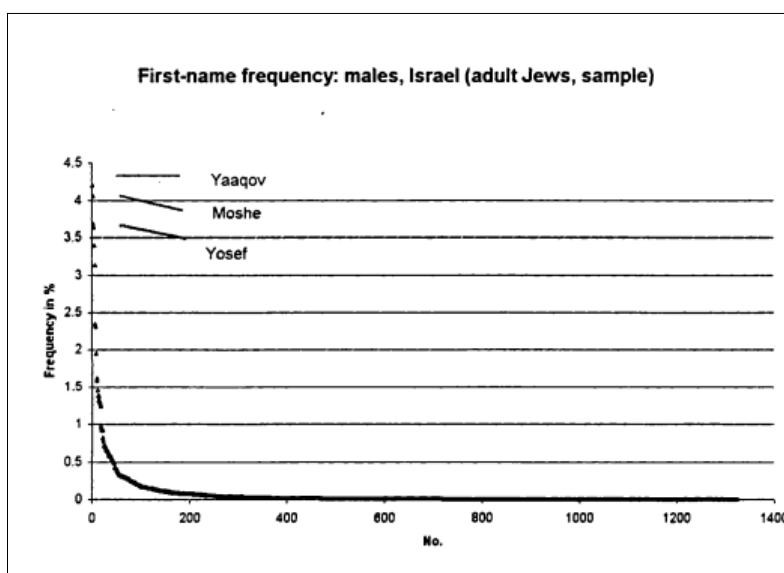
²⁵ W anglojęzycznych pracach słowa *type* (typ) używa się na określenie jednostki leksykalnej, zaś *token* (okaz) – jej wystąpienie w tekście. Zbiór typów tworzy słownik wyrazów użytych w tekście, natomiast zbiór okazów odpowiada długości tekstu, sumie użyć wszystkich wyrazów (Sambor 1972: 219, Tucker 2001: 70).

ułożonej malejąco pod względem częstości występowania. Pierwsze 10 imion, stanowiących niecałe 0,8% słownika imion, wyczerpuje aż połowę wszystkich użyć, zaś 50 najczęstszych antroponimów (niecałe 4% słownika) porywa już 90% całego zbioru.

Z. Abramowicz (1993: 371), badając imiona białostoczan, wyodrębniła z kolei trzy zauważalne klasy częstości imiennictwa, z których każda spełnia określoną funkcję – imiona o wysokiej w pewnych okresach frekwencji względnej (wyższą niż 2% sumy użyć), imiona średnio popularne ($1\% > 2\%$), imiona bardzo rzadkie ($< 0,5\%$). Dwie pierwsze grupy, niezbyt liczne, stanowią zdaniem autorki podstawę współczesnego systemu imienniczego Polaków.

Omówione zjawisko nierównomiernego rozkładu imion w zbiorach, choć powszechnie obserwowane, nie doczekało się w polskiej literaturze dotyczącej imiennictwa szerszego omówienia. Zagadnieniem tym bardziej szczegółowo zajęli się natomiast badacze zagraniczni. A. Eshel (2001), badając żydowskie imiona męskie nadawane w Izraelu sporządził wykres empirycznego rozkładu częstości imion, ukazującego zależność między rangą (oś x) a frekwencją względną imienia (oś y), który cechował się skrajną asymetrią prawostronną (rys. 6), co jest charakterystyczne dla danych językowych (Hammerl, Sambor 1990: 35). Wykres ten pokazuje, jak wraz ze wzrostem rangi imienia maleje jego frekwencja. Dla porównania Eshel sporządził analogiczne wykresy dla imion nadawanych w Stanach

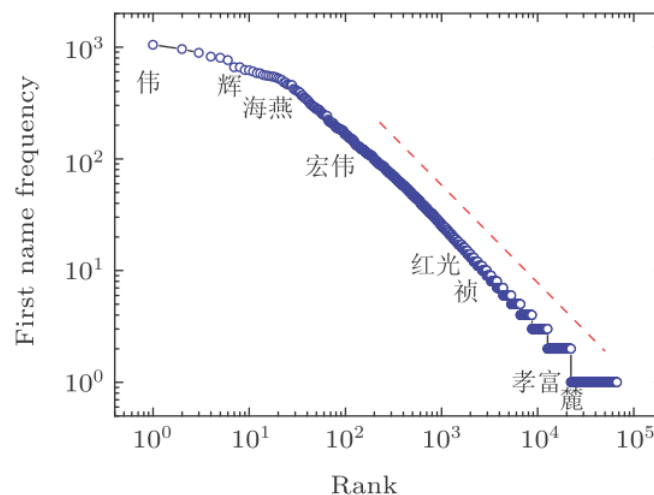
Rys. 6. Wykres zależności między rangą (oś x) i frekwencją w izraelskim imiennictwie (za: Eshel 2001: 56)



Zjednoczonych i Berlinie oraz wszystkich imion w Izraelu (zbiory te znacząco różniły się pod względem wielkości), które przybrały podobny kształt, jednak wystąpiły między nimi pewne różnice, polegające m.in. na różnym stopniu pokrycia sumy użyć przez r najczęstszych imion. Według autora miały one swoje pozajęzykowe przyczyny związane z różną strukturą etniczną

poszczególnych obszarów. Na tej podstawie Eshel postulował stosowanie graficznego obrazu rozkładów częstości w postaci wykresów w celu analizy różnic w imiennictwie różnych krajów, a także między imiennictwem jednego kraju w różnych punktach czasowych lub w obrębie różnych grup społecznych (Eshel 2001: 59). W tym samym czasie ukazały się artykuły D. A. Galbiego (2001) i D. K. Tuckera (2001), którzy sporządzili wykresy zależności między rangą a częstością imion nadawanych w USA, stosując skalę podwójnie logarytmiczną. Wykazali oni, że w obrębie określonego przedziału rang punkty odpowiadające kolejnym imionom z listy frekwencyjnej mają przebieg liniowy, co oznacza, że między rangą a frekwencją zachodzi relacja potęgowa. Potwierdziły to również późniejsze prace M. W. Hahna i R. A. Bentleya (2003), istnienie tej zależności ujawniły też badania na materiale imion nadawanych Kanadzie i Wielkiej Brytanii (Tucker 2002, 2004) oraz w Chinach (Guo i in. 2011), co obrazuje wykres na rys. 7.

Rys. 7. Wykres zależności ranga-częstość w imiennictwie chińskim w skali logarytmicznej dla losowej próby ok. 220 tys. osób (za: Guo i in. 2011: 4)

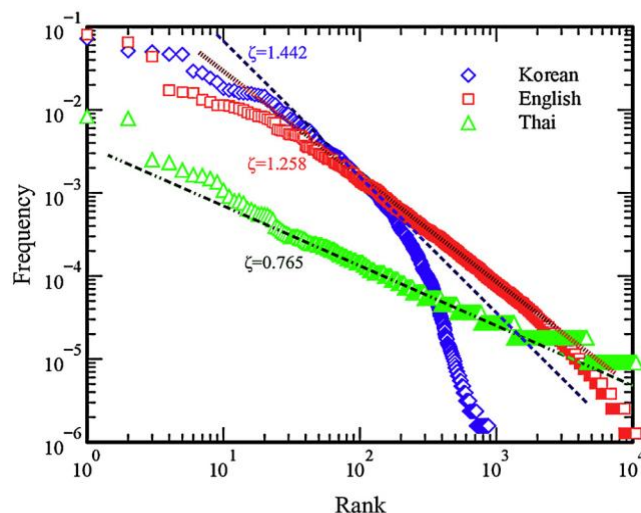


Zależność ta, w językoznawstwie znana pod postacią prawa Zipfa, charakterystyczna jest dla tekstów języka naturalnego. Wyrażana jest uogólnionym²⁶ wzorem $f_r \propto r^{-\alpha}$, gdzie f_r to częstość wyrazu o danej randze, r to ranga imienia na liście frekwencyjnej, zaś wykładnik potęgowy α to współczynnik kątowy funkcji liniowej, według której przebiegają punkty w pewnym zakresie rang (im większy kąt padania linii prostej w stosunku do osi x, tym wyższa jest wartość α). W części wykresu, w której punkty przebiegają wzdłuż linii prostej

²⁶ W różnych tekstach można się spotkać z odmienną notacją. Stosowane w dalszej części określenie parametr α jako wskaźnik ilościowego bogactwa słownictwa tekstu u J. Sambor (1969: 39) występuje jako parametr B (oznaczenie pochodzące z modyfikacji Mandelbrota).

wykładnik α ma wartość stałą. W oryginalnym ujęciu model G. K. Zipfa zakładał, że niezależnie od języka, autora czy epoki wartość parametru α dla całego tekstu jest stała i wynosi 1, jednak liczne późniejsze badania wykazały, że zależność liniowa nie zachodzi w obszarze najwyższych i najniższych rang²⁷, zaś sama wartość α na ogół jest wyższa od 1. W językoznawstwie porównawczym, zwłaszcza w badaniach stylistycznych, prawo Zipfa jest ważne przede wszystkim właśnie ze względu na parametr α – może on służyć jako miernik różnic w ilościowej strukturze słownictwa tekstu, można go też interpretować jako wskaźnik jego ilościowego bogactwa: im większa wartość α , tym uboższe ilościowo słownictwo (Sambor 1972: 68–71). Przykładem takiego zastosowania jest opublikowane niedawno badanie porównawcze przekładów Biblii w stu różnych językach autorstwa A. Mehri i M. Jamaati (2017) – na rys. 8 zamieszczono wykres przedstawiający przebieg krzywych ranga- częstość wraz z prostymi dopasowania (w obszarze rang spełniających prawo Zipfa) z odpowiadającymi im wartościami α (tu oznaczonego jako ζ) w tekście koreańskim, angielskim i tajskim.

Rys. 8. Porównanie wykresów ranga- częstość dla przekładów Biblii w językach koreańskim, angielskim i tajskim (za: Mehri, Jaamati 2017: 2472)



Parametr α z równania Zipfa wykorzystywany jako wskaźnik bogactwa leksykalnego tekstu stosowany jest także w badaniach porównawczych imiennictwa w różnych punktach czasowych i na różnych obszarach. Jego zmienność w czasie w obrębie imion męskich i żeńskich nadawanych w kolejnych latach z przedziału 1880–2010 w Stanach Zjednoczonych przedstawił W. Li (2012: 46–49). Do lat 20. wyższe wartości α (u Li jest to parametr a)

²⁷ Widoczne jest to na wykresie na rys. 7 – w obszarze pierwszych, wysokich rang obserwujemy pewne zagięcie, zaś w przypadku rang najwyższych, przypisanych imionom najrzadszym mamy do czynienia z charakterystycznymi „schodkami”.

notowane były dla imion żeńskich, w kolejnych dwóch dekadach nastąpiło pod tym względem zrównanie imiennictwa żeńskiego i męskiego, zaś od lat 40. większe wartości α występują już konsekwentnie w obrębie imion męskich, co świadczyć może o ich mniejszym bogactwie ilościowym w stosunku do imion żeńskich. W pierwszych dekadach rozpatrywanego okresu wartość parametru α w przypadku obu grup wykazywała umiarkowaną tendencję spadkową, następnie do lat 60. obserwujemy znaczny wzrost wartości badanego wskaźnika, po czym widoczny jest jej regularny spadek aż do roku 2010. Generalnie parametr α w badaniach Li przyjmował wartości między 1,3 a 1,9.

Podobne badania na materiale amerykańskim przeprowadziła grupa krakowskich badaczy (Pomorski i in. 2016: 172–174). Zbadali oni nie tylko zmienność wartości wykładnika α w następujących po sobie latach z przedziału 1910–2011 (również i tu zaobserwowano zmianę trendu w latach 60. oraz wartości α podobne do tych z badania Li), ale dokonali także porównania między grupami stanów powiązanych ze sobą historycznie i kulturowo, co ujawniło znaczne różnice między nimi. Stany południowe wykazywały się najwyższymi wartościami α , największa różnorodność dotyczyła zaś grupy stanów zachodnich. Można też zauważyć tendencję polegającą na stopniowym zmniejszaniu się różnic pomiędzy tymi regionami od lat 60. Autorzy dokonali także porównań wartości parametru α obliczonego dla imion nadawanych w Belgii (osobno dla imion męskich i żeńskich) w jej trzech głównych regionach: Flandrii, Walonii i Brukseli w podziale na trzy grupy wiekowe (< 18, 18–64, > 65 lat). Również i tu ujawniły się pewne różnice regionalne i czasowe, które można ująć jako następujące prawidłowości: 1) wraz ze wzrostem wieku zwiększa się wartość parametru α , a więc imiennictwo w najmłodszej grupie wiekowej było najbardziej różnorodne (niezależnie od regionu); 2) wartości α dla stołecznej Brukseli były znacznie niższe niż w pozostałych regionach. Nie odnotowano natomiast żadnych wyraźnych różnic między imiennictwem męskim i żeńskim pod tym względem.

Wykładnik α z równania Zipfa nie jest oczywiście jedynym, ani też najczęściej stosowanym parametrem rozkładu częstości, którym posługiwano się w celu zbadania zmienności struktury ilościowej imiennictwa między różnymi jego zbiorami. Pokrycie sumy użyć przez określoną liczbę najczęstszych imion, odzwierciedlające u S. Gajdy (1973) stopień konkretyzacji upodobań rodziców, było przedmiotem szczegółowych analiz we wspomnianych już pracach S. Liebersona (2000: 42–68) oraz S. Liebersona i F. Lynn (2003: 238–243), którzy traktują ten wskaźnik jako miarę koncentracji imiennictwa. Na materiale imion nadawanych na przestrzeni co najmniej 200 ostatnich lat w Stanach Zjednoczonych,

Kanady, Anglii, Walii, Szkocji, Francji, Niemczech, Danii oraz na Węgrzech w różnych przedziałach czasowych, autorzy przebadali procentowe pokrycie globalnej sumy użyć przez najczęstszych 10, 15, 20 lub 25 imion²⁸, co doprowadziło ich do wniosku, że na przełomie XIX i XX wieku nastąpił znaczny i postępujący w następnych latach spadek koncentracji w imiennictwie wszystkich badanych krajów. Szczegółowe badania tego typu przeprowadzili również na materiale amerykańskim przywołani w poprzednim rozdziale Twenge i in. (2010, 2016), którzy szczegółowo zmierzili zmienność w czasie procentowego udziału 1, 10, 25 i 50 najczęstszych imion w globalnej sumie użyć w skali ogólnamerykańskiej oraz w poszczególnych stanach. Wyniki ich obliczeń naniesione na wykresy pokazały, że imiennictwo męskie zawsze było bardziej skoncentrowane niż żeńskie, a także potwierdziły odnotowany przez Liebersona generalny trend malejącej koncentracji, jednak zauważono, że stałość tego trendu zależna jest od liczby dobieranych najczęstszych imion. Wskazują na to wykonane dodatkowo obliczenia współczynnika korelacji Pearsona między wskaźnikami procentowego udziału a następującymi po sobie latami. Przykładowo w obrębie imion męskich dla jednego najczęstszego imienia ujemna korelacja jest niemal pełna ($\rho = -0,94$), co oznacza, że z każdym kolejnym rokiem koncentracja malała, jednak w przypadku 10, 25 i 50 imion jest to już odpowiednio: -0,84, -0,73, -0,64 (Twenge i in. 2010: 21). Pokazuje to, że stosowanie różnej liczby najczęstszych imion do zmierzenia w ten sposób stopnia koncentracji może dać odmienne wyniki. Niezależnie od tej kwestii, w każdym przypadku zdecydowanie większy udział najczęstszych imion notowany był w obrębie imiennictwa męskiego.

Wspomniano już wcześniej, że Twenge i in. coraz mniejszy udział najczęstszych imion w globalnej sumie użyć interpretowali jako przejaw tendencji indywidualistycznej w zachowaniach nazewniczych, które najsilniej według nich zaznaczyły się w USA w pod koniec XX w. Odmienne kwestię tę postrzega J. Gerhards (2005: 67–70), który symptomów indywidualizacji szuka nie w klasie imion najczęstszych, lecz najrzadszych. To właśnie chęć nadania oryginalnego imienia jest zdaniem badacza tym, co najistotniej świadczy o indywidualistycznym charakterze wyboru rodzica. Za miarę obrazującą stopień indywidualizacji badacz uznał udział w ogólnej sumie użyć imion nadanych tylko jednej osobie – antroponimicznych hapakslegomenów. W badanych przez niego niemieckich miejscowościach zaobserwowano gwałtowny wzrost udziału hapakslegomenów w pierwszej połowie XX w. – od ok. 40% na przełomie wieków do niemal 80% na początku lat 1950,

²⁸ Z nieznanymi mi powodów w różnych krajach badano udział różnej liczby najczęstszych imion w sumie użyć.

odkąd udział ten utrzymywał się już na stałym poziomie. Jak już wspomniałem wcześniej, również S. Gajda (1973) uwzględniał najrzadszą leksykę antroponimiczną w swoim badaniu porównawczym dla Opola i okolic, jednak otrzymanych wyników nie poddał interpretacji.

W. Li (2012: 45) zwrócił uwagę, że mierzenie nierównomierności rozkładu w imiennictwie wyłącznie poprzez obliczenie udziału w nim określonej liczby najczęstszych lub najrzadszych jednostek jest niewystarczające. Do tego celu zdecydowanie bardziej odpowiedni jest znany z nauk ekonomicznych wskaźnik Giniego²⁹, w badaniach językoznawczych częściej określany mianem wskaźnika Lorenza, przy pomocy którego określa się właśnie stopień skoncentrowania tekstu (Sambor 1969: 158, Skowronek 2001: 83). Wskaźnik ten mierzy pokrycie danego tekstu (w naszym przypadku – zbioru imion) przez częstość kolejnych wyrazów (imion) na liście rangowej, a więc bierze pod uwagę zniuansowanie rozkładu wszystkich jej elementów, a nie tylko np. imion najczęstszych. Wskaźnik ten może przyjmować wartości od 0 (w przypadku całkowicie równomiernego rozkładu) do 1 (przy pełnej koncentracji).

Obliczone przez Li wartości wskaźnika Giniego dla imion męskich i żeńskich w USA w kolejnych latach z przedziału 1880–2010 pokazały nieco odmienny trend w koncentracji niż przedstawiony przez Liebersona – w pierwszych dekadach XX w. stopień koncentracji ulegał zwiększeniu, następnie utrzymywał się na względnie stałym, wysokim poziomie, by dopiero po roku 1980 ulegać stopniowemu zmniejszaniu.³⁰ Generalnie jednak koncentracja w imiennictwie amerykańskim jest bardzo wysoka (pomiędzy 0,8 na początku badanego okresu a niemal 0,95 w latach 50.). Wyniki uzyskane przez Li łączy natomiast z omówionymi wcześniej badaniami to, że mniejsze nierówności, niższa koncentracja (a więc większa różnorodność) występują w obrębie imiennictwa żeńskiego (Li 2012: 46).

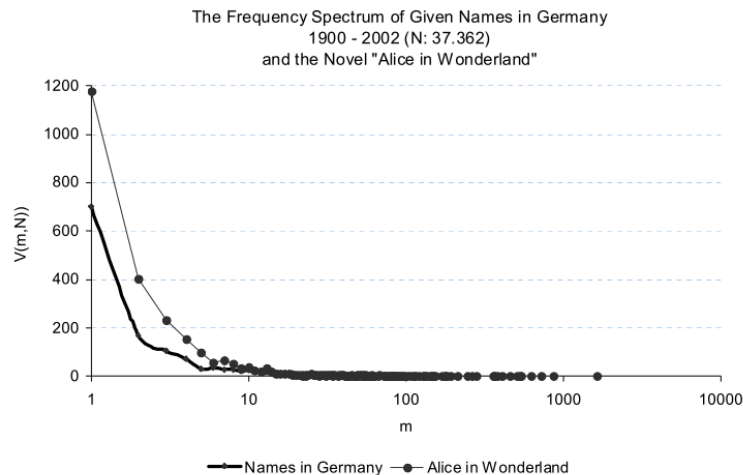
Niemieccy badacze D. Huschka i G. Wagner w publikacji *Statistical Problems and Solutions in Onomastic Research* (2010), odwołując się do badań ze statystyki leksykalnej, poruszyli bardzo ważną kwestię równoliczności zbiorów przy porównywaniu ich rozkładów częstości w imiennictwie oraz bazujących na nich charakterystyk liczbowych. Autorzy również zwrócili uwagę na podobieństwo dystrybucji częstości wyrazów w zbiorach imion i tekstach języka naturalnego, co obrazują wykresy spektrum częstości (ang. *frequency*

²⁹ Wskaźnik Giniego w ekonomii nazywany jest również wskaźnikiem nierówności społecznej – pierwotnie stosowano go przy opisie nierównomierności rozkładu dochodów w społeczeństwie.

³⁰ Badacz zakłada, że spadek koncentracji może być powiązany ze wzmożonym ruchem imigracyjnym, wzbogacenie amerykańskiego systemu imienniczego o nowe jednostki.

spectrum), czyli „ślepych” list frekwencyjnych³¹ dla imion nadanych w Niemczech w latach 1900–2002 oraz tekstu powieści *Alicja w krainie czarów* L. Carolla (rys. 9).

Rys. 9. Porównanie wykresów spektrów częstości zbioru imion nadawanych w Niemczech w latach 1900–2002 i tekstu *Alicji w krainie czarów* L. Carolla (za: Huschka, Wagner 2010: 16)



Zbiory imion, tak jak teksty języka naturalnego, charakteryzują się typem dystrybucji, określanym w statystyce jako *Large Number of Rare Events* (LNRE), cechującym się występowaniem wielu elementów o niskiej frekwencji. Ich liczba (a więc ostatecznie również i liczba wszystkich elementów w zbiorze) jest bardzo podatna na zmiany wielkości zbioru. Oznacza to, że wszelkie statystyki opisowe bazujące na różnych obszarach list frekwencyjnych również będą wrażliwe na wielkość badanego zbioru. Badacze przekonują, że jeśli chcemy przy ich pomocy porównywać różne zbiory imion, to podobnie jak w przypadku badań z zakresu statystyki leksykalnej, musimy operować na próbach o takiej samej lub zbliżonej wielkości, gdyż w praktyce nie istnieją parametry rozkładu częstości (np. różne wskaźniki bogactwa ilościowego) odporne na długość tekstu (w naszym przypadku wielkość zbioru imion).

Autorzy krytycznie odnieśli się do wspomnianych wcześniej prac Eshela czy Tuckera, którzy dokonywali porównań zbiorów o diametralnie różnej wielkości³² oraz przedstawili metody sporządzania liczbowo równych prób na potrzeby badań porównawczych.

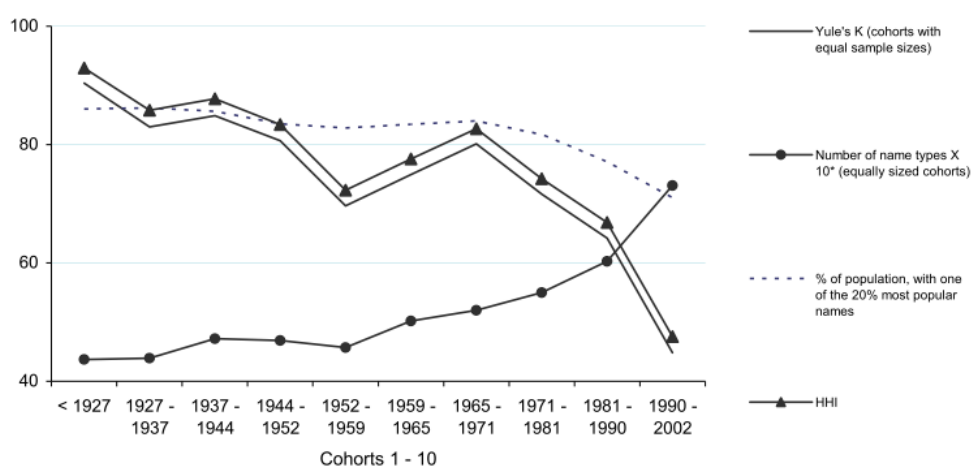
³¹ W odróżnieniu od zwykłej, pełnej listy frekwencyjnej, zawierającej spis wszystkich wyrazów według malejącego porządku częstości oraz odpowiadające im częstości f oraz rangi r , spektrum częstości, czyli „ślepa lista” zawiera tylko częstości f_i oraz odpowiadające im liczby wyrazów w_i bez wyszczególnienia, o które wyrazy chodzi (Sambor 1969: 29).

³² Przykładowo Eshel porównywał pod kątem procentowego pokrycia globalnej sumy użyć przez r najczęstszych imion liczący 632 tys. zbior imion męskich z Berlina (bazujący na książkach telefonicznych) ze zbiorem imion męskich nadanych w USA, który obejmował 134 mln osób.

O zagadnieniu równości prób szerzej piszę w rozdziałach III.1 i III.2. W tym miejscu należy jeszcze wspomnieć, że Huschka i Wagner, poza używanymi w dotychczasowych badaniach porównawczych imiennictwa wskaźnikami bogactwa ilościowego, jak liczba typów czy pokrycie sumy użyć przez r najczęstszych imion, zaproponowali również stosowanie znanego z badań z zakresu statystyki leksykalnej wskaźnika różnorodności K Yule'a i bardzo zbliżonego do niego wskaźnika koncentracji Herfindahla-Hirschmanna (HHI) stosowanego w naukach ekonomicznych³³.

Podobnie jak w przypadku wspomnianego wcześniej wskaźnika Giniego, autorzy podkreślali fakt, iż obejmują one pełny rozkład częstości, a nie tylko sferę obejmującą imiona najpopularniejsze. Zwrócili też uwagę, że porównań imiennictwa różnych zbiorów najlepiej dokonać przy pomocy większej liczby wskaźników ilościowych, a nie tylko jednego. Przeprowadzone przez nich badania na materiale imienniczym z Niemiec, podzielonym na dziesięć równych pod względem liczby osób przedziałów z okresu 1900–2002, pokazały, że na jego przestrzeni znacząco powiększył się słownik imion używanych w Niemczech przy jednoczesnym zmniejszeniu się koncentracji w imiennictwie, również świadczącym o wzroście jego różnorodności (zob. rys. 10). Badania ujawniły też, że imiennictwo żeńskie odznacza się bogatszym słownikiem i niższą koncentracją niż imiona męskie (podział na płeć nie został uwzględniony na wykresie).

Rys. 10. Porównanie wybranych wskaźników bogactwa leksykalnego w zbiorach imion nadawanych w 10 przedziałach czasowych w Niemczech (za: Huschka, Wagner 2010: 30)



³³ W ekonomii służy on do określenia poziomu zagęszczenia w danej branży oraz poziomu konkurencji na danym rynku.

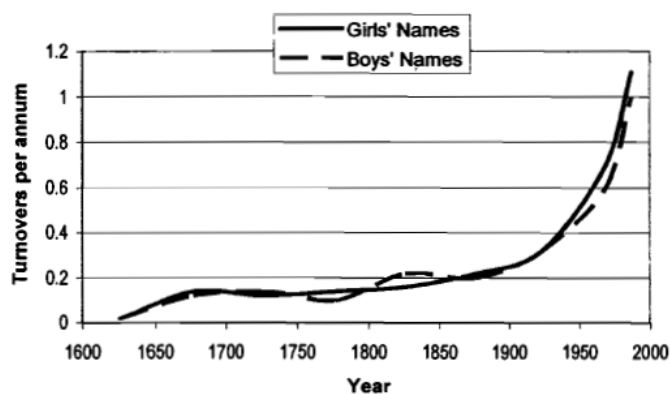
Zwróćmy uwagę, że na powyższym wykresie przebieg krzywych dla wartości poszczególnych wskaźników nie jest w pełni skorelowany – wzrostowi liczby typów nie zawsze towarzyszy spadek koncentracji, również trendy w koncentracji mierzonej różnymi wskaźnikami są nieco odmienne. Przemawia to tym bardziej za tym, że w badaniach porównawczych imiennictwa, tak jak zresztą w przypadku komparatystycznych badań z zakresu statystyki leksykalnej czy stylometrii, warto stosować możliwie wiele wskaźników bogactwa ilościowego, gdyż opieranie się tylko na jednym z nich może doprowadzić nas do zbyt pochopnych wniosków.

Zjawiska związane z nierównomiernością rozkładu częstości imion w różnych zbiorach, które opisuje się przy pomocy wymienionych wskaźników to obszerna, choć nie jedyna kwestia, jaka poruszana jest przy badaniu diachronicznego i synchronicznego zróżnicowania systemów imienniczych. W badaniach dotyczących dynamiki zmian w imiennictwie jedną z często poruszanych kwestii jest tempo, w jakim w kolejnych punktach czasowych do zasobu najczęściej używanych imion wchodzi nowe jednostki antroponimiczne, wypierając starsze, przy czym nie chodzi tu o nowo utworzone imiona, ale takie, które dotychczas w tym zasobie nie występowały. Jak już wspomniałem, w pracach S. Gajdy (1973) i E. Umińskiej (1983) tempo tej wymiany mierzone było poprzez obliczenie liczby zgodności pomiędzy listami 10 najczęstszych imion w kolejnych punktach czasowych, natomiast w większości późniejszych prac zagranicznych badaczy, opisujących bardziej długofalowe zjawiska w imiennictwie (m.in. Wilson 1998, Lieberson 2000, Hahn, Bentley 2003, Gureckis, Goldstone 2009) używana jest miara odwrotna, mianowicie liczba różnic pomiędzy listami najczęstszych imion, którą określa się mianem wskaźnika tempa rotacji (ang. *turnover rate*). Zazwyczaj oblicza się go na podstawie liczby różnic pomiędzy listami najczęstszych imion w następujących po sobie interwałach czasowych (np. 5- czy 10- letnich) podzielonej przez liczbę lat z danego przedziału. Posługując się tym prostym wskaźnikiem obliczanym dla 10, 20 czy 25 więcej najczęstszych imion, wspomniani wcześniej amerykańscy badacze S. Lieberson i F. Lynn, wykazali, że w imiennictwie USA, Kanady i siedmiu krajów europejskich na przestrzeni XX w. doszło do niespotykanego w poprzednich stuleciach przyspieszenia i wzrostu w zakresie wymiany najpopularniejszych imion w kolejnych sekwencjach czasowych (zob. rys. 13), spowodowanego działaniem mechanizmów mody (Lieberson 2000: 36–66, Lieberson, Lynn 2003: 235, 255; Lieberson, Lynn 2008: 139–141). Do podobnych wniosków co Lieberson doszli także inni badacze imiennictwa – w przypadku Francji i Wielkiej Brytanii S. Wilson (1998: 317–318, 326), Niemiec – J. Gerhards (2005:

94), Stanów Zjednoczonych – T. M. Gureckis, R. L. Goldstone (2009: 654–655)³⁴. Nieco odmienne wyniki dla Stanów Zjednoczonych w XX w. otrzymali M. W. Hahn i R. A. Bentley (2003: 121–122), którzy co prawda również odnotowali zwiększenie tempa rotacji³⁵ na przestrzeni całego wieku, ale proces ten nie przebiegał u nich regularnie (zob. rys. 11). Naukowcy ci jednak w swoich badaniach opierali się na symulacji, generując wielokrotnie losowe próby (na bazie 5% danych z całego okresu) dla każdego przedziału czasowego. Wyniki badań Hahna i Bentleya zgodne są natomiast z obserwacjami Gureckisa i Goldstone’a (2009: 654) co do różnego stopnia tempa rotacji wśród imion nadawanych kobietom i mężczyznom.

Obie te prace pokazują, że od początku XX wieku wskaźnik rotacji wśród imion żeńskich jest wyraźnie wyższy niż wśród imion męskich, co oznacza, że prawdopodobieństwo nadania nowego imienia jest większe w przypadku dziewczynek niż chłopców³⁶. Hahn i Bentley przyczyn tego zjawiska upatrują m.in. w patriarchalnym modelu kultury, jednak nie bez znaczenia jest też fakt, iż zasób imion biblijnych, które stanowią jedno z podstawowych źródeł współczesnych zachodnich systemów imienniczych, w ogromnej mierze tworzą imiona męskie, podczas gdy zaledwie 5,5–8% imion biblijnych to imiona żeńskie (Bohmbach 2000: 33–34), co w pewnym sensie wymusza uzupełnianie zasobu imion żeńskich o nowe jednostki innego pochodzenia (Hahn, Bentley 2003: 122).

Rys. 11. Liczba nowych pozycji wśród 20 najczęstszych imion w Danii od XVII do XX w. (za: Lieberson, Lynn 2003: 236)

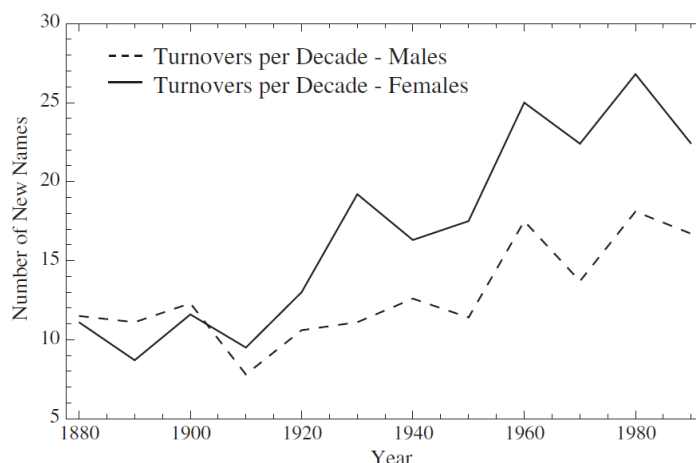


³⁴ Gureckis i Goldstone badali tempo rotacji na 1000 najpopularniejszych imion w każdym roku (zob. rys. 12).

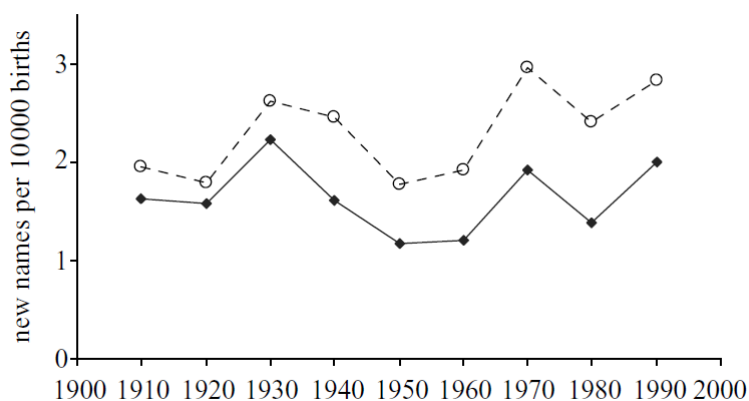
³⁵ Na określenie zjawiska innowacji imienniczej autorzy używają zapożyczonego z genetyki terminu *tempo mutacji* (ang. *mutation rate*), przeciwstawnego względem replikacji.

³⁶ Z wliczeń Hahna i Bentleya wynika, że stosunek średniej tempa rotacji dla poszczególnych dekad (od 1919 do 1990 r.) wynosi 2,3 : 1,6 nowych imion odpowiednio dla kobiet i mężczyzn na 10 tys. urodzeń (test *t* dla zmiennych powiązanych, $t = 6,59$, $p = 0,0002$). Oznacza to, że prawdopodobieństwo otrzymania przez osobę płci żeńskiej nowego imienia wynosi 1,4.

Rys. 12. Liczba nowych pozycji wśród 1000 najpopularniejszych imion w kolejnych dekadach w USA (za: Gureckis, Goldstone 2009: 654)



Rys. 13. Liczba nowych pozycji wśród 1000 najpopularniejszych imion w kolejnych dekadach w USA z uwzględnieniem różnic w wielkości populacji (linia przerywana – im. żeńskie, linia ciągła – im. męskie) (za: Hahn, Bentley 2003: 122)



W powyższych przykładach obliczenie liczby różnic (lub zgodności) między listami rangowymi r najczęstszych imion w kolejnych punktach czasowych służyło ukazaniu dynamiki zmian w obrębie najczęstszego imiennictwa w perspektywie czasowej, diachronicznej. W ten sam prosty sposób można też określać „bliskość” lub „oddalenie” (używając terminów B. Hryniewicz-Adamskich) kilku (pod)systemów imienniczych w perspektywie synchronicznej. Przykład takiego rozwiązania znajdziemy w artykule S. Gajdy, w którym autor zestawiał liczby zgodności między listami dziesięciu najczęstszych imion nadawanych na obszarze powiatu namysłowskiego, oleskiego oraz miasta Opola (w podziale na ludność rodzimą i napływową). Porównania dokonano dla każdej pary list – największa zbieżność (tj. największa liczba wspólnych imion) dotyczyła obszarów jednorodnych pod względem pochodzenia regionalnego mieszkańców i odwrotnie – najmniej zgodności wystąpiło między listami sporządzonymi dla obszarów zamieszkałych z jednej strony przez ludność autochtoniczną, z drugiej zaś – napływową (Gajda 1973: 281).

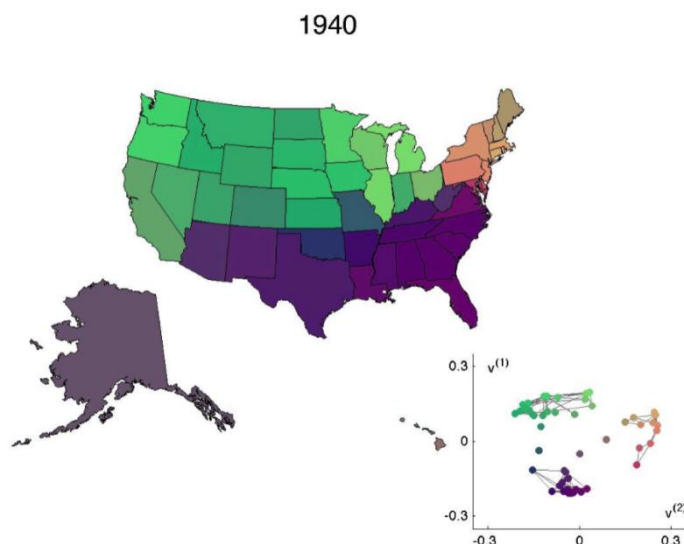
Nie zawsze jednak wskazanie części wspólnych i rozłącznych badanych zbiorów imienniczych (zwłaszcza gdy mowa tylko o niewielkiej liczbie najczęstszych pozycji) jest wystarczające dla ukazania relacji między nimi i konieczne jest sięgnięcie do informacji o frekwencji i randze poszczególnych imion. Listy najczęstszych imion nie muszą bowiem różnić się wcale (lub mogą różnić się w bardzo nikłym stopniu) pod względem jednostek antroponimicznych, które się na nie składają, ale już udział tych konkretnych jednostek w ogólnej sumie użyć (na podstawie którego przypisywana jest im ranga) może stanowić istotny czynnik dyferencyjny. P. Swoboda (2013: 56–67), porównując rangi, jakie w poszczególnych województwach w latach 1995–2010 zajmowały najczęstsze imiona w skali ogólnopolskiej, wykazał, że istnieją pewne zwarte obszary geograficzne, charakteryzujące się częstszym występowaniem jakiegoś imienia bądź grupy imion. Główny nacisk został tu jednak położony na przedstawienie imiennictwa typowego dla danego obszaru, wykazującego wyraźne odchylenie od sytuacji ogólnopolskiej³⁷ (czyli w pewnym sensie uśrednionej). Przykładowo imiona *Edyta* i *Elżbieta* najniższe rangi miały w województwach podlaskim, małopolskim i podkarpackim, więc możemy powiedzieć, że pod tym względem województwa te są podobne, bo wymienione imiona są dla nich typowe. Jednak już imię *Elwira*, choć najbardziej popularne było również w woj. podlaskim i podkarpackim, to stosunkowo rzadko, także na tle innych województw, występowało w małopolskim, zaś imię *Janusz* z kolei rzadziej pojawiało się w podlaskim. Takie żmudne zestawienia można przeprowadzić dla wszystkich imion, które występują w badanych zbiorach, ale pokaże to jedynie, że pod kątem częstości występowania określonej grupy imion niektóre zbiory są do siebie podobne, natomiast nie pozwoli przedstawić całej złożonej sieci powiązań, jakie zachodzą między nimi, wynikających często z subtelnych różnic w dystrybucji wielu imion w obrębie poszczególnych list frekwencyjnych.

W przywoływanych już w poprzednim rozdziale dwóch pracach dotyczących regionalnego zróżnicowania imiennictwa w USA (Barucca i in. 2015, Pomorski i in. 2016) autorzy zaprezentowali rozwiązania metodologiczne, które pozwalają kompleksowo zbadać i spojrzeć z większego dystansu na te sieci powiązań bez przyglądania się dystrybucji geograficznej każdego z imion z osobna. W obu pracach wykorzystuje się zbliżone, choć w szczegółach nieco odmienne statystyczne metody grupowania, które bazują na analizie korelacji między

³⁷ Można to porównać do określania stopnia dyspersji wyrazów w tekstach reprezentujących poszczególne style funkcjonalne języka, mającego na celu oddzielenie zasobu słownictwa neutralnego stylistycznie, wspólnego dla wszystkich stylów (wspólnostylowego, ponadstylowego) od słownictwa nacechowanego, typowego dla poszczególnych stylów (Kamińska-Szmaj 1990: 5).

rozkładami częstości imion w poszczególnych stanach. Siła korelacji między dwiema zmiennymi jest tym wyższa, im bardziej podobny jest układ imion na listach frekwencyjnych. Badacze włoscy obliczyli częstości względne wszystkich imion nadanych w każdym ze stanów i na tej podstawie ustalili współczynniki korelacji liniowej Pearsona dla każdej pary stanów (obliczenia te osobno wykonano dla każdego roku z przedziału 1910–2012). Wyniki obliczeń zebrali w macierz korelacji, odzwierciedlającą siłę powiązań między każdą parą stanów w poszczególnych latach. Na podstawie tych danych autorzy ukazali, w jaki sposób poszczególne stany łączą się ze sobą w większe zgrupowania (klastry) oraz jaki jest stopień relacji podobieństwa pomiędzy nimi. W tym celu zastosowali niezależnie dwie zaawansowane procedury statystyczne z zakresu analizy czynnikowej – analizy głównych składowych (ang. principal component analysis, PCA) oraz hierarchicznej analizy skupień (hierarchical cluster analysis, HCA)³⁸, których zadaniem jest wykrycie pewnych struktur wynikających ze związków między zmiennymi i klasyfikacja obiektów w obrębie tych struktur. Na rys. 14 zamieszczono graficzny obraz analizy głównych składowych przeprowadzonej na materiale imion żeńskich nadawanych w poszczególnych stanach w roku 1940.

Rys. 14. Grupowanie stanów USA metodą analizy głównych składowych w oparciu o podobieństwo rozkładu częstości imion na listach frekwencyjnych (za: Barucca i in. 2015: 7945)

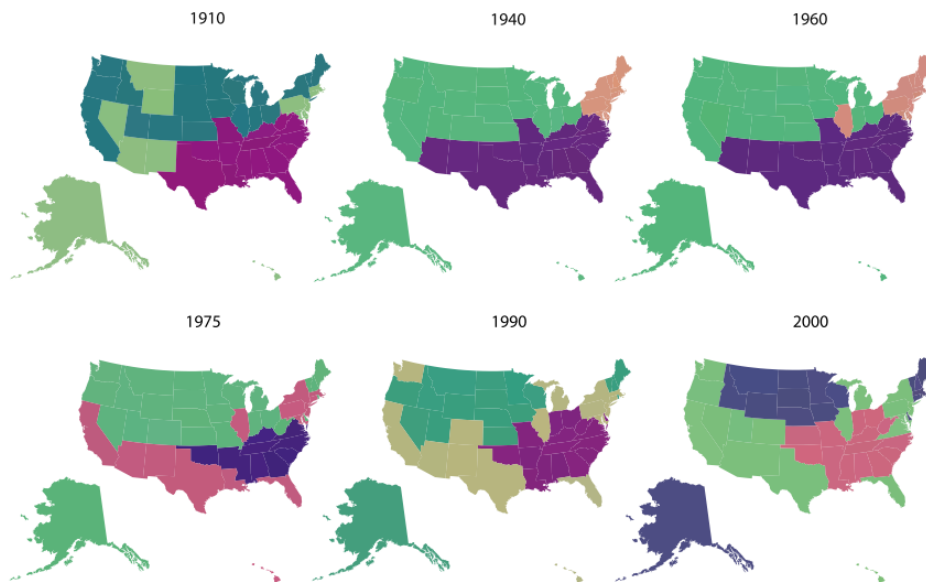


Obszary, które zaznaczono podobnym kolorem, wykazywały większą zbieżność w zakresie rozkładu częstości imion. Różnice w zakresie imiennictwa pomiędzy różnymi

³⁸ Metodę hierarchicznej analizy skupień, którą zastosowałem w niniejszej pracy, szczegółowo opisuję w rozdziałach V.1–4.

stanami odzwierciedlają odległości pomiędzy poszczególnymi punktami na biplocie (dwuwykresie) w dolnej części rysunku. Przeprowadzenie tego typu analiz na zbiorach imion nadawanych w kolejnych latach pozwoliło autorom zaobserwować przeobrażenia tych relacji w czasie, odzwierciedlonych w układzie geograficznych konfiguracji. Na rys. 15 przedstawiono z kolei kartograficzny obraz hierarchicznej analizy skupień, który ukazuje rozmieszczenie geograficzne w kilku punktach czasowych trzech najwyrazistszych grup stanów (każdej grupie przyporządkowano inny kolor) w obrębie których odnotowano najsilniejsze korelacje.

Rys. 15. Rozmieszczenie geograficzne trzech największych klastrów stanów uzyskanych w wyniku hierarchicznej analizy skupień (za: Barucca i in. 2015: 7946)

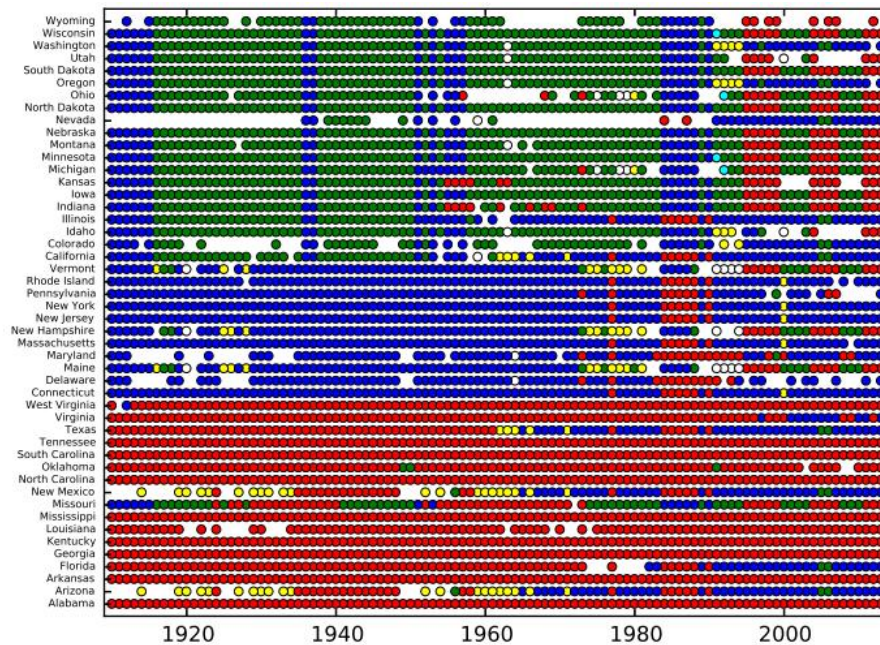


Grupa krakowskich naukowców (Pomorski i in. 2016) do badania zróżnicowania geograficznego imiennictwa Stanów Zjednoczonych wykorzystała zaczerpnięte z nauk społecznych narzędzia analizy sieciowej³⁹ (ang. network analysis). Podobnie jak we wcześniej omówionej pracy, podstawą analizy była tu macierz korelacji Pearsonowskich między każdą parą stanów, jednak ich obliczenia bazowały na frekwencji względnej tylko stu najczęstszych imion w całych USA nadawanych w kolejnych latach z przedziału 1910–2010 w każdym ze stanów. Macierz ta stanowi swego rodzaju sieć, którą tworzą badane jednostki (w tym przypadku poszczególne stany reprezentowane przez zbiory imion), zwane węzłami

³⁹ Określanej też mianem społecznej analizy sieciowej. Metodę tę wykorzystuję pomocniczo w części podsumowującej wyniki analizy zróżnicowania geograficznego w zakresie parametrów rozkładu częstości (III.4).

lub też punktami. Powiązane są one ogniwami, o których sile decyduje w tym przypadku siła korelacji. Celem analizy sieciowej jest odkrycie pewnych wzorów struktur czy też społeczności (ang. *communities*) w obrębie sieci i ich wzajemnych relacji. Wyniki przeprowadzonego badania (przedstawione na rys. 16), mimo że opierało się ono na nieco okrojonych zmiennych, zredukowanych jak już wspomniano do 100 najpopularniejszych imion, okazały się niemal identyczne z zaprezentowanymi przez badaczy włoskich.

Rys. 16. Ewolucja w czasie struktury społeczności sieci powiązań stanów w oparciu o korelacje między listami 100 najczęstszych imion (za: Pomorski i in. 2016: 172)



Warto wspomnieć, że podobnie jak wiele omówionych wcześniej metod porównywania zbiorów imienniczych przy pomocy parametrów rozkładu częstości, również opisane powyżej metody, takie jak analiza głównych składowych, hierarchiczna analiza skupień czy analiza sieciowa, są od lat z powodzeniem stosowane w badaniach językoznawczych, przede wszystkim z zakresu stylometrii.⁴⁰

Z powyższego przeglądu publikacji wynika, że w badaniach nad imiennictwem w coraz większym stopniu (zwłaszcza w ostatnich latach) przywiązuje się wagę do właściwości jego struktury statystycznej. Przedstawione prace pokazują, że charakterystyki liczbowe zbiorów imion mogą stanowić ich cechy dystynktywne w badaniach porównawczych systemów

⁴⁰ Zob. np. liczne publikacje autorstwa osób skupionych w grupie Computational Stylistics Group związanej z Instytutem Języka Polskiego PAN (<https://computationalstylistics.github.io/publications/>).

imienniczych w perspektywie czasowej, przestrzennej czy społecznej. Niektóre z zaproponowanych w nich metod (wraz z innymi metodami zaczerpniętymi z pola statystyki leksykalnej) zostały wykorzystane również w niniejszej pracy w celu zbadania zróżnicowania w obrębie polskiego imiennictwa.

II.2. Założenia metodologiczne

Głównym celem analiz przeprowadzonych w rozdziałach III, IV i V jest zbadanie przy pomocy odpowiednich metod statystycznych wewnętrznego zróżnicowania systemu imienniczego w Polsce przez pryzmat cech ilościowych, charakteryzujących zbiory imion nadawanych w różnych wycinkach czasowych oraz na różnych obszarach. Badanie to pozwoli nam sprawdzić, czy czasowa i przestrzenna konfiguracja wyodrębnionych zbiorów imion przejawia jakąś tendencję, która może odzwierciedlać uwarunkowania pozajęzykowe związane z określonymi wydarzeniami historycznymi (kształtującymi procesy kulturowo-społeczne) lub też określonym typem regionalizacji obszaru współczesnej Polski, które możemy ująć jako czasową i terytorialną stratyfikację społeczeństwa polskiego.

Do przeprowadzenia badania użyto m.in. metod z zakresu językoznawstwa statystycznego, przede wszystkim statystyki leksykalnej i stylometrii. Jednym z założeń lingwistyki kwantytatywnej jest przekonanie, że frekwencja poszczególnych elementów w języku i jej rozkład może być cechą dystynktywną np. idiolektu, języka grupy społecznej, określonego gatunku tekstu bądź stylu danego autora (Sambor 1972: 13, 19–20, Dunaj 1988: 83, Eder 2014: 92–95). Antroponimy, jak też inne nazwy własne, choć pod wieloma względami (gramatyki, funkcji, pragmatyki czy ortografii) odróżniają się od sfery apelatywnej, to bez wątplenia należą do zasobu leksykalnego języka (Malec 2007: 125), co pozwala nam zakładać, że możemy badać je przy użyciu narzędzi z zakresu statystyki leksykalnej. Zbiory imion nadanych przez rodziców w różnym czasie i miejscu, podobnie jak np. zbiory wyrazów użytych przez autorów różnych tekstów, mogą się bowiem różnić między sobą nie tylko cechami jakościowymi. Traktowane jako pewne struktury ilościowe (w których każdemu elementowi odpowiada dana częstość użycia) będą różniły się także cechami wyrażanymi wyłącznie liczbowo, nie odnoszącymi się wprost do formy językowej imion. Cechą taką może być na przykład liczba różnych (antropo)leksemów użytych w każdym ze zbiorów, czyli wielkość ich słownika, lub ich układ na liście rangowej. Takie ujęcie – potraktowanie zbioru nazw osobowych na podobnych zasadach jak tekstu w celu ich zbadania przy pomocy metod kwantytatywnych nie jest na gruncie antroponomastyki nowe. Zastosowała je w pracy dotyczącej struktury ilościowej nazwisk polskich K. Skowronek, proponując termin *tekst onomastyczny* na określenie danego zbioru antroponimów (Skowronek 2001: 26–27). Z przeglądu prac w rozdziale II.3 również wynika, że metody znane z badań statystyczno-leksykalnych są stosowane także w badaniach nad imiennictwem.

Zanim przejdę do omówienia narzędzi i metod statystycznych wykorzystanych przeze mnie w analizie, muszę jednak w pierwszej kolejności ustalić kategorie socjolingwistyczne, stanowiące kryteria wyodrębniania z materiału badawczego poszczególnych zbiorów imion. Przyjmując bowiem za cel badania wykrycie zależności między wewnętrznym zróżnicowaniem systemu imienniczego a czasową i terytorialną stratyfikacją społeczeństwa, w którym on funkcjonuje, opieram się na założeniu, że imię jest nie tylko zjawiskiem językowym, ale także zjawiskiem socjalnym. Rozwój i funkcjonowanie imion są ściśle związane ze społeczeństwem ludzkim, jego rozwojem, historią i kulturą – kształt systemu imienniczego i jego zmienność motywowane są tak czynnikami lingwistycznymi, jak i ekstralingwistycznymi, społecznymi (Abramowicz 1994: 152). Tak jak poszczególne teksty literackie mają swoich autorów, a odmiany lub style funkcjonalne języka (reprezentowane przez teksty) mają swoich użytkowników, również i zbiory imion nadawanych w określonym miejscu i czasie, swoiste „teksty onomastyczne”, mogą być kształtowane przez cechy ich twórców funkcjonujących w określonym kontekście społecznym. D. Bartol-Jarosińska podaje za W. Labovem, że „zewnętrzne przyczyny wariantywności wywodzą się «z charakterystyki mówiącego i charakteru używania języka». Na charakterystykę mówiącego składają się cechy statusowe, które dają się ująć jako czynniki demograficzne, klasowo-warstwowe, środowiskowe” (Bartol-Jarosińska 1991: 57). Cechy te na wyższym poziomie ogólności możemy sprowadzić do trzech podstawowych wymiarów: terytorialnego (przestrzennego), społecznego (warstwowego) i czasowego (Piotrowski, Ziółkowski 1976: 115). Stosując te uwarunkowania do problematyki poruszanej w niniejszej pracy, a także do charakteru materiału badawczego, za „mówiącego” musimy uznać nie tyle pojedyncze osoby (np. rodzica), które dokonują aktu nominacji, ale całą społeczność językową czy też wspólnotę komunikatywną, która korzysta z określonego repertuaru językowego (zob. Bokszański i in. 1977: 58), ściślej – imienniczego. Informacje zawarte w materiale badawczym pozyskanym z bazy PESEL pozwalają nam na wyznaczenie wewnętrznych granic społeczności językowej poprzez dwie kategorie: przynależności terytorialnej (osoby dokonujące aktu nominacji np. w danym miejscu lub regionie) i/lub przynależności czasowej (osoby dokonujące aktu nominacji w danym roku lub przedziale czasowym), stanowiące – obok innych kategorii – społeczne osie wariantywności języka.⁴¹ O „charakterze używania języka” stanowią z kolei

⁴¹ D. Bartol-Jarosińska (1991: 58–59) do osi społecznej wariantywności zalicza następujące kategorie: wykształcenie, zawód, przynależność terytorialną, pochodzenie, wiek i płeć. Informacje dotyczące imion, które uzyskaliśmy z bazy PESEL, ograniczają się do imienia, roku i miejsca rejestracji urodzenia oraz płci osoby otrzymującej imię. Na tej podstawie bezpośrednio możemy określić jedynie przynależność terytorialną osób

wybrane przez te społeczności środki językowe, w tym przypadku nadane przez nie imiona tworzące cały zbiór antroponimów, które składają się na ich repertuar imienniczy, opisywany przy pomocy odpowiednich charakterystyk liczbowych.

Poniżej w sposób bardziej szczegółowy określone zostały wspomniane kategorie przynależności czasowej i terytorialnej – owe osie wariantowości systemu imienniczego, według których zostały wyodrębnione poszczególne zbiory imion.

Materiał badawczy analizowany w niniejszej pracy analizuję w trzech perspektywach czasowo-przestrzennych, przy czym w każdej z nich osobno rozpatruję i porównuję zbiory imion żeńskich i męskich, co ma na celu ukazanie ewentualnych różnic w zachowaniach nazewniczych warunkowanych czynnikiem płci osoby otrzymującej imię. W pierwszej perspektywie celem badania jest ilościowe zróżnicowanie w czasie imiennictwa ogólnopolskiego. Wydzielone zostały z niego możliwie najmniejsze podzbiory odpowiadające kolejnym następującym po sobie latom z badanego okresu. Nie czynię tu zatem żadnych założeń periodyzacyjnych. Przeciwnie – analiza ilościowa repertuarów imienniczych stanowiących efekt kolektywnych zachowań ogółu osób nadających imiona w poszczególnych latach w Polsce (a więc swego rodzaju społeczności językowej określonej przez przynależność czasową) ma tu charakter *stricte* eksploracyjny. Ma ona na celu sprawdzenie czy na przestrzeni badanego okresu w określonych punktach lub przedziałach czasowych ujawniły się wyraźne tendencje (które w określonych przypadkach mogą świadczyć o funkcjonowaniu swego rodzaju czasowej odmiany ilościowej imiennictwa). Natomiast na poziomie interpretacji tych zjawisk można odnieść się do ustaleń z zakresu periodyzacji historyczno-językowej. Ramy czasowe, w jakich osadzony jest materiał badawczy, obejmują bowiem szereg wydarzeń, swego rodzaju punktów zwrotnych, których efektem były rozmaite zmiany kulturowe, społeczne, polityczne czy demograficzne, rezonujące wyraźnie także w języku. Znajduje to wyraźne odbicie w charakterze wyodrębnianych przez historyków języka okresów w najnowszych dziejach polszczyzny: okres sprzed odzyskania przez Polskę niepodległości (do 1918 r.), okres międzywojenny (1918–1939), II wojna światowa (1939–1945) oraz okres powojenny, dzielący się na kilka krótszych okresów: migracji i interferencji (1945–1960), kształtowania się wspólnoty komunikatywnej (1960–1980), trwania i dezintegracji stereotypów (1980–1990), okres przewartościowań (1990–) (Klemensiewicz 1981: 517–527, Bajerowa (red.) 1996: 437–439,

dokonujących aktu nominacji, zaś kategorię wieku, o którym nie jesteśmy w stanie nic orzec w odniesieniu do tych osób, na potrzeby pracy spróbujmy zastąpić kategorią przynależności czasowej.

Dubisz 1995: 83–85). Przemiany w strukturze ilościowej imiennictwa również mogą w jakimś stopniu odzwierciedlać ten typ periodyzacji, choć oczywiście musimy pamiętać, że w niniejszej pracy badane są tylko pojedyncze aspekty jego funkcjonowania.

Druga z perspektyw dotyczy przestrzennego zróżnicowania ilościowego polskiego imiennictwa, choć również z uwzględnieniem czynnika czasu⁴². Badanie w takim ujęciu ma na celu sprawdzenie, czy w zakresie cech ilościowych możemy mówić o swego rodzaju regionalnych odmianach polskiego systemu imienniczego oraz czy w kolejnych sekwencjach czasu przestrzenna wariancja ilościowa imiennictwa ulegała przemianom. Do jego przeprowadzenia konieczne jest wydzielenie z materiału badawczego zbiorów imion pod kątem przynależności terytorialnej uwzględnionych w nim osób. Zdefiniowanie przestrzennego zakresu takich regionalnych wspólnot jest sprawą dosyć kłopotliwą i wymaga generalizacji. Istnieje bowiem wiele typologii regionów współczesnego obszaru Polski, które uwzględniają rozmaite kryteria – historyczne, geograficzne, etniczne, językowe, gospodarcze, kulturowe, administracyjne itd. Regiony różnego typu, choć wyodrębniane na podstawie zobiektywizowanych kryteriów, są zawsze sztucznymi konstruktami tworzonymi przez specjalistów w celu uporządkowania i klasyfikowania rzeczywistości (Jałowiecki 1996: 41). Granice tych regionów mogą (choć nie muszą) się wzajemnie pokrywać, a ich przebieg często jest płynny i różnie definiowany. Zazwyczaj uważa się, że najbardziej trwałe i ciągle widoczny w wielu dziedzinach życia społecznego typ podziału regionalnego współczesnej Polski pokrywa się z granicami dawnych państw zaborczych⁴³ oraz Ziemi Zachodnich i Północnych, uzyskanych po 1945 r. Powszechnie też obserwuje się podział Polski na wschód i zachód (tzw. „pęknięcie” na linii Wisły). Wydzielone tak obszary cechuje odmienność struktur demograficznych, społecznych, gospodarczych, zachowań oraz postaw społecznych i politycznych, a także krajobrazu historyczno-kulturowego (Jałowiecki 1996: 6, 21–25, 36, Bartkowski 2003: 132, Gorzelak 2007: 196–198, Plit 2015: 87–88). W niniejszej pracy nie zakładam z góry żadnego modelu konfiguracji przestrzennej imiennictwa i postanowiłem – z dużą dozą arbitralności – w pierwszej kolejności zbadać zróżnicowanie imiennictwa pomiędzy obszarami współczesnych szesnastu województw. Ich granice administracyjne oczywiście tylko w pewnym stopniu pokrywają się z regionami

⁴² Z uwagi na sygnalizowane już luki materiałowe zróżnicowanie badanie przestrzennego zróżnicowania imiennictwa w większości przypadków ograniczam do okresu po 1945 r.

⁴³ Wliczając w to obszary takie jak Górny Śląsk, który był częścią Prus na długo przed rozbiorem Polski, zamieszkały był jednak w znacznej mierze przez polską ludność autochtoniczną.

historyczno-kulturowymi⁴⁴, jednak w celu stwierdzenia, czy istnieje pewne zróżnicowanie przestrzenne imiennictwa w ogóle oraz czy jego kształt odzwierciedla w przybliżeniu dany typ regionalizacji, jest to podział wystarczający. Jest to też najwygodniejszy sposób podziału materiału badawczego, którym dysponujemy. W poszczególnych przypadkach, w celu weryfikacji, czy zaobserwowana w toku analizy ilościowej imiennictwa konfiguracja województw przybiera pewien charakterystyczny kształt, możemy oczywiście dokonać bardziej szczegółowej analizy na poziomie powiatów czy gmin, których granice łatwiej można dopasować do granic np. regionów historycznych⁴⁵.

Trzecia perspektywa zakłada porównanie pod kątem cech ilościowych imiennictwa czterech wybranych ośrodków wielkomiejskich (Wrocławia, Krakowa, Warszawy i Poznania) na tle otaczających ich obszarów o niższym stopniu urbanizacji (również w sekwencjach czasowych). Można podejrzewać bowiem, że czynniki kulturowo-społeczne związane z zamieszkaniem w różniących się pod względem stopnia urbanizacji miejscach mogą odgrywać rolę w poziomie zmienności zachowań ich mieszkańców, także językowych, do których niewątpliwie zalicza się nadawanie imion. Powszechnie uważa się, że mieszkańcy miast cechują się większym liberalizmem w zakresie wyborów, stylu życia czy stosunku do tradycji niż ma to miejsce w przypadku mieszkańców tzw. prowincji, którym przypisuje się raczej postawę zachowawczą (CBOS 2014: 8). Jak pisze J. Warchala:

„Zakres i tempo zmian językowych zależne jest przede wszystkim od czynnika mobilności społecznej. Im bardziej osiadłe i «nieruchome» społeczeństwo, pozostające w zamkniętych terytorialnie homogenicznych grupach, tym mniejsze tempo zmian w systemie gramatycznym i leksykalnym języka. Wytwarza się wówczas społeczny układ konserwujący stan języka na danym terytorium (...), a który zmniejsza szanse wymiany informacji, a w rezultacie interferencji języków” (Warchala 2003: 153).

Według K. Skowronek (2013: 124–125) „zróżnicowanie mieszkańców wielkich miast (ze względu na ich pochodzenie etniczne, wykształcenie, zawód, poglądy polityczne, społeczne

⁴⁴ Przykładowo współczesne woj. śląskie składa się z historycznych ziem Górnego Śląska i Małopolski, uwzględniając zaś granice polityczne sprzed 1918 r., na jego obszar składają się tereny należące do Prus, Królestwa Kongresowego i Austro-Węgier. W okresie międzywojennym podzielony był z kolei między Niemcy i Polskę (tu dodatkowo między woj. kieleckie i krakowskie oraz autonomiczne woj. śląskie). Innym przykładem jest woj. mazowieckie, które obejmuje swoim zasięgiem zarówno historyczne Mazowsze, jak i północną Małopolskę. Na woj. wielkopolskie składają się z kolei obszary tak zaboru pruskiego, jak i rosyjskiego (kaliskie), zaś w okresie międzywojennym jego niewielka północno-zachodnia część należała do Niemiec.

⁴⁵ Tak jest w przypadku rozdziału IV.5, w którym przyglądam się zagadnieniu rotacji imiennictwa na obszarze niemieckich stref okupacyjnych z okresu II wojny światowej, oraz rozdziału V. 6, gdzie analizowane jest zróżnicowanie imiennictwa w woj. śląskim w kontekście przynależności jego poszczególnych obszarów do dawnych mocarstw zaborczych.

i religijne) powoduje, że tu mają swój początek wszelkie „mody” nazewnicze, „fale” popularności różnych imion etc.”, możemy zatem wstępnie założyć, że także w zakresie cech ilościowych, np. wskaźników bogactwa leksykalnego czy dynamiki zmian w imiennictwie ośrodki wielkomięskie będą się wyróżniały na tle obszarów prowincjonalnych. Z drugiej strony należy jednak pamiętać, że krajobraz społeczny polskich miast i obszarów wiejskich na przestrzeni badanego okresu ulegał ciągłym przeobrażeniom warunkowanym czynnikami politycznymi, ekonomicznymi czy cywilizacyjnymi. Dotyczy to nie tylko samego procesu urbanizacji, która w pierwszej połowie XX w. była na stosunkowo niskim poziomie – w miastach mieszkało wówczas poniżej 30% mieszkańców Polski (Skowronek 2013: 99), by dopiero II wojnie światowej osiągnąć znaczny wzrost – od 42,5% w roku 1950 do 61,8% w roku 2000 (Szymańska 2002: 54). O ile bowiem w pierwszej połowie XX w. możemy mówić o wykluczeniu z obiegu kultury znacznej części społeczeństwa żyjącej poza miastami, to wraz z postępującym w późniejszych latach rozwojem cywilizacyjnym czy technologicznym oraz tendencjami globalistycznymi nastąpiła częściowa niwelacja różnic w zakresie stylu życia, a także zachowań językowych między mieszkańcami miast i prowincji (Warchała 2003: 155–161).

Wyodrębnione w taki sposób repertuary imiennicze, czyli zbiory imion nadawanych w określonym czasie i na określonym terytorium, traktowane przeze mnie jako „teksty onomastyczne”, w kolejnych rozdziałach zostaną poddane analizie przy użyciu wybranych metod z zakresu statystyki leksykalnej, a także niektórych metod zaprezentowanych w rozdziale II.1.3, używanych w analizach porównawczych imiennictwa. Z uwagi na różnorodność zastosowanych metod ograniczę się w tym miejscu jedynie do ich zwięzłego zestawienia, natomiast ich bardziej szczegółowy opis znajduje się w odpowiednich rozdziałach.

W rozdziale III dokonano analizy porównawczej w zakresie realizacji wybranych parametrów rozkładu częstości wyrazów w tekście na podstawie list frekwencyjnych sporządzonych dla zbiorów imion wyodrębnionych w ramach każdej w trzech perspektywach czasowo-przestrzennych: w imiennictwie ogólnopolskim w kolejnych latach z przedziału 1906–2010, w imiennictwie na obszarze współczesnych województw w sześciu następujących po sobie dekadach w latach 1951–2010, w imiennictwie wybranych miast na tle województw, w których się znajdują również w sześciu następujących po sobie dekadach w latach 1951–2010. Do parametrów tych należą następujące wskaźniki bogactwa leksykalnego: wielkość słownika (W), udział w imiennictwie najrzadszej leksyki (hapaks-,

dis- i trislegomenów), wskaźnik oryginalności Guirauda, wskaźnik różnorodności Yule'a (K), wskaźnik stereotypowości Mistríka, parametr α z prawa Zipfa oraz wskaźniki koncentracji – Lorenza, Mistríka oraz stosowany w niektórych pracach jako miara koncentracji procentowy udział r najczęstszych imion w sumie użyć wszystkich antropoleksemów w danym zbiorze. Zasadnicza część analityczna (III.3) poprzedzona została jednak rozdziałami, w których opisane zostało zagadnienie równoliczności porównywanych zbiorów oraz procedury skalowania korpusów imienniczych, koniecznego przy badaniach porównawczych.

W rozdziale IV również we wszystkich trzech przedstawionych perspektywach przeprowadzono analizę dynamiki zmian w obrębie najczęstszych imion. Wykorzystano w tym celu stosowany w badaniach nad imiennictwem wskaźnik tempa rotacji, mierzący liczbę zmian, jakie zachodziły w obrębie najczęstszego imiennictwa w kolejnych sekwencjach czasu (5- i 10-letnich). Również w tym badaniu uwzględniono wszystkie trzy wymienione perspektywy czasowo-przestrzenne. W ostatnim podrozdziale (IV.4) przeprowadzono dodatkowo szczegółowe badanie tempa rotacji w skali rocznej, które ujawniło bardzo wyraźne „zdarzenia” w tym zakresie w okresie II wojny światowej, zasługujące na odrębne omówienie.

W rozdziale V w celu zbadania regionalnego zróżnicowania imiennictwa w zakresie układu imion na listach frekwencyjnych na poszczególnych obszarach wykorzystano metodę hierarchicznej analizy skupień. Metoda ta, szeroko wykorzystywana w badaniach językoznawczych, zwłaszcza stylometrycznych, należy do działu wielowymiarowej analizy statystycznej. Jej głównym celem jest ułożenie w grupy obiektów (w tym przypadku zbiorów imion nadawanych na określonym terytorium), reprezentowanych przez określone cechy (w tym przypadku częstość względną poszczególnych imion w tych zbiorach) w taki sposób, aby odzwierciedlić stopień ich powiązania. Pozwala ona na wyłonienie spośród badanych zbiorów imienniczych najistotniejszych skupień (klastrów), które cechują się największym podobieństwem ze względu na frekwencję względną poszczególnych imion. Badanie zostało przeprowadzone na poziomie województw w kolejnych latach z okresu 1946–2010 oraz powiatów województwa śląskiego w dwudziestoletnich przedziałach z okresu 1931–2010.

Wszystkie czynności związane z porządkowaniem i analizą materiału badawczego, obliczeniami oraz wizualizacją wyników w postaci wykresów, kartogramów i dendrogramów wykonałem przy użyciu języka programowania i środowiska do obliczeń statystycznych *R* (R Core Team 2015), wykorzystując graficzny interfejs użytkownika RStudio. Poza jego natywnymi funkcjami w dużej mierze korzystałem z pakietów narzędzi z zakresu analizy

lingwistycznej stworzonych w środowisku R: *stylo* (Eder i in. 2013), *languageR* (Baayen 2013) oraz *zipfR* (Evert, Baroni 2007). Do stworzenia kartogramów wykorzystałem narzędzia z zakresu przestrzennej analizy danych z pakietu *maptools* (Bivand, Levin-Koh 2016), natomiast przy tworzeniu dendrogramów, będących wynikiem hierarchicznej analizy skupień, posłużyłem się narzędziami z pakietu *dendextend* (Galili 2015).

III. Parametry rozkładu częstości wyrazów w zbiorach imienniczych

W niniejszym rozdziale dokonano porównawczej analizy zbiorów imienniczych pod kątem realizacji wybranych wskaźników bogactwa leksykalnego w polskim imiennictwie, na które składają się: wielkość słownika (W), udział w imiennictwie najrzadszej leksyki (hapaks-, dis- i trislegomenów), wskaźnik oryginalności Guirauda, wskaźnik różnorodności K Yule'a, wskaźnik stereotypowości Mistríka, parametr α z prawa Zipfa oraz miary koncentracji – Lorenza, Mistríka, a także zaproponowany przez S. Liebersona jako miara koncentracji procentowy udział r najczęstszych imion w sumie użyc wszystkich antropoleksemów w danym zbiorze⁴⁶. Niektóre z wymienionych wskaźników – wielkość słownika, wskaźniki oryginalności, różnorodności, stereotypowości oraz miary koncentracji Lorenza i Mistríka – były już przeze mnie używane do wstępnej oceny regionalnego zróżnicowania ilościowego imiennictwa w Polsce (Swoboda 2015). Są to klasyczne miary bogactwa leksykalnego powszechnie stosowane w pracach porównawczych z zakresu statystyki leksykalnej (por. np. Sambor 1969, Kamińska-Szmaj 1990, Gąsiorek 1997, Lewandowska 2008). W niniejszej pracy częściowo korzystam z wyników tego badania, postanowiłem jednak wykorzystać kilka dodatkowych miar, takich jak udział najrzadszej leksyki (hapaks-, dis- i trislegomenów), parametr α oraz procentowy udział najczęstszej leksyki, z uwagi na ich stosowanie w pracach z zakresu antroponimii (por. m.in. Lieberson 2000, Li 2012, Pomorski i in. 2016). Lista parametrów rozkładu częstości używanych do określania bogactwa leksykalnego tekstów jest znacznie dłuższa (ich obszerny przegląd znajdziemy m.in. w pracach: Tweedie, Baayen 1998, Pawłowski 2003, Tuldava 2005), jednak część z nich opiera się na bardzo podobnych założeniach i nie ma sensu ich równoległe stosowanie, niektóre zaś znajdują zastosowanie głównie w badaniach nad tekstami literackimi, gdyż odnoszą się do ich struktury kompozycyjnej. Sposób obliczania i interpretacji użytych przeze mnie wskaźników został przedstawiony w poszczególnych poświęconych im podrozdziałach.

Z uwagi na znany fakt, że właściwie wszystkie charakterystyki liczbowe tekstu zależne są od jego długości N (Tweedie, Baayen 1998: 323 i n., Baayen 2001: 24 i n., Pawłowski 2003: 175), w badaniach porównawczych konieczne jest posługiwanie się zbiorami imion o takiej samej wielkości. Sporządzenie list frekwencyjnych i obliczenie na ich podstawie wartości wymienionych wskaźników wymagało zatem uprzedniego zastosowania procedur mających na celu wyrównanie ich pod względem długości. Opis procedur skalowania korpusów

⁴⁶ Wskaźnik ten celowo analizuję obok pozostałych wskaźników koncentracji w celu ich łatwiejszego porównania.

imienniczych, jak i samo zagadnienie równości prób zostały przedstawione w rozdziałach III.1 i III.2.

W poszczególnych podrozdziałach części III.3. dokonałem zwięzłego omówienia wymienionych wskaźników oraz ich obliczeń i analizy osobno dla imion żeńskich i męskich w kilku perspektywach czasowych i przestrzennych: A) w imiennictwie ogólnopolskim w kolejnych latach z przedziału 1906–2010, B) w imiennictwie na obszarze współczesnych województw w sześciu następujących po sobie dekadach w latach 1951–2010, C) w imiennictwie wybranych miast na tle województw, w których się znajdują – również w sześciu następujących po sobie dekadach w latach 1951–2010.

Wyniki analizy starałem się przedstawić w sposób ilustracyjny i możliwie zwięzły. W przypadku badania A z uwagi na dużą rozpiętość czasową wyniki przedstawiono na wykresach, których oś x stanowi linia czasu (lata z 1906–2010), natomiast na osi y znajdują się wartości danego wskaźnika. Na wykres naniesiono krzywe dla imion nadanych kobietom i mężczyznom, które ukazują zmienność w czasie wartości danych wskaźników. Dodatkowo sporządzono wykresy słupkowe, na których przedstawiono poziom różnic w wartościach wskaźników między imiennictwem żeńskim i męskim w poszczególnych latach. Poziom ten to zwykła różnica – od wartości dla imion żeńskich odjęto te dla męskich – jeśli wynik jest dodatni, świadczy to o wyższych wskazaniach w imiennictwie żeńskim. Pozwoli to nam prześledzić, jak kształtowały się w czasie wartości poszczególnych statystyk, czy wystąpiły między nimi zbieżności, a także czy występują różnice dotyczące struktury liczbowej pomiędzy imiennictwem męskim i żeńskim.

W badaniu B, dotyczącym ilościowego zróżnicowania geograficznego imiennictwa, w tabelach zamieszczono wartości poszczególnych statystyk obliczone dla imion z obszaru każdego z 16 współczesnych województw⁴⁷ w każdej z sześciu wspomnianych dekad. W dwóch ostatnich wierszach tabeli zamieszczono średnią wartość danego wskaźnika oraz współczynnik zmienności v , którym posłużyłem się w celu określenia zróżnicowania w czasie wartości wskaźników między województwami.⁴⁸ Dane umieszczone w tabeli dodatkowo

⁴⁷ W tym i w kolejnych rozdziałach pracy w tabelach i na części rysunków poszczególne województwa oznaczane są następującymi skrótami: DŚL – dolnośląskie, KUJ. – kujawsko-pomorskie, LBL. – lubelskie, LBS – lubuskie, ŁDZ. – łódzkie, MŁP. – małopolskie, MAZ. – mazowieckie, OPL. – opolskie, PDK. – podkarpackie, PDL. – podlaskie, POM. – pomorskie, ŚLS. – śląskie, ŚWK. – świętokrzyskie, WRM. – warmińsko-mazurskie, WLP. – wielkopolskie, ZPM. – zachodniopomorskie.

⁴⁸ Współczynnik zmienności (v) ujmuje stosunek odchylenia standardowego do średniej arytmetycznej, dla przejrzystości zwykle mnożony jest przez 100. Współczynnik ten pozwala na ocenę zmienności cechy, abstrahując od jej jednostek (Hammerl, Sambor 1990: 54). Wyższa wartość v świadczy o większym zróżnicowaniu między poszczególnymi obiektami (w naszym przypadku – województwami).

naniesiono także na mapę administracyjną Polski (każdej dekadzie odpowiada jeden kartogram), co ułatwi ocenę tego, czy w zakresie danego wskaźnika ujawnia się trwały układ danych zależny od czynników przestrzennych, czy też ma on charakter przypadkowy. Skala odcieni szarości w legendzie mapy opiera się na rozpiętości wartości badanej statystyki we wszystkich sześciu dekadach.

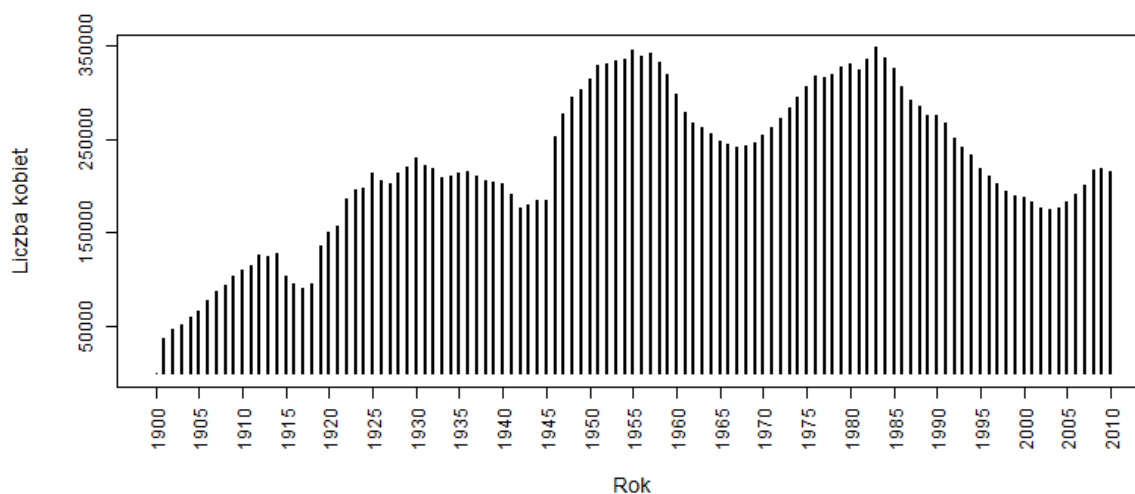
W części C dokonano porównawczej analizy wskaźników bogactwa leksykalnego między czterema ośrodkami wielkomiejskimi – Krakowem, Poznaniem, Warszawą i Wrocławiem – a obszarem województw, których się znajdują – małopolskiego, wielkopolskiego, mazowieckiego i dolnośląskiego. Podobnie jak w badaniu B również i tu operuję sześcioma dekadami. W kolejnych wierszach tabeli zamieszczono wartości wskaźników obliczonych dla miast oraz dla województw (obliczone z wyłączeniem materiału z wymienionych miast) oraz wartość różnicy między tymi statystykami, pełniącej tu podobną funkcję do współczynnika zmienności z badania B.

III.1. Zagadnienie wielkości i równości korpusów imienniczych

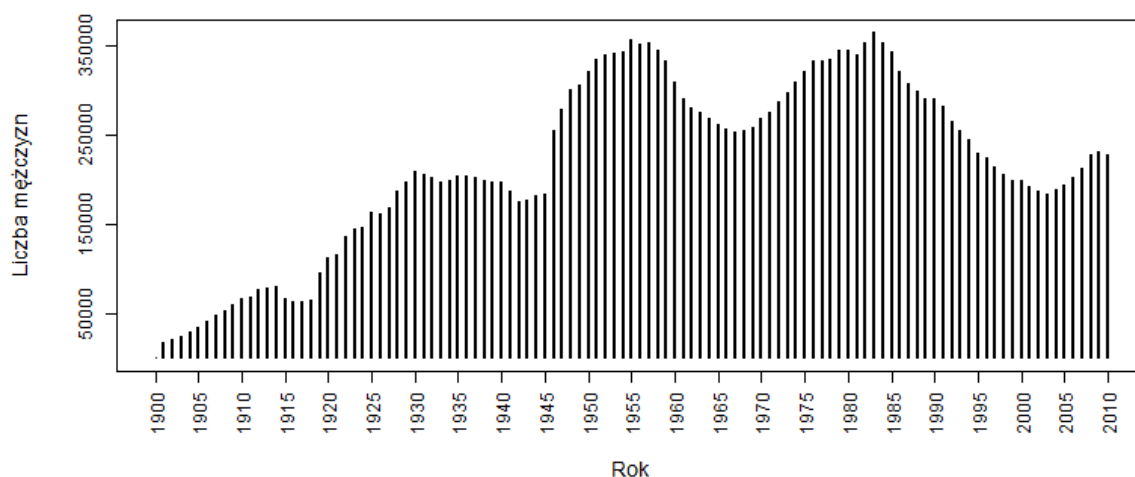
Przy analizie materiału imienniczego pochodzącego z bazy PESEL (ale też z innych źródeł) pod kątem wskaźników bogactwa leksykalnego, tak w aspekcie diachronicznym, jak i synchronicznym, należy mieć na uwadze pewną, wydawałoby się, oczywistą kwestię. Jest on mianowicie ilościowo niejednorodny. W każdym z analizowanych lat liczba osób, którym nadano imię była różna. System PESEL został w pełni wdrożony w 1984 r. i dopiero od tego roku zaczęto ewidencjonować w nim wszystkie urodzone w Polsce osoby. Natomiast jeśli chodzi o osoby urodzone przed tym rokiem, to w rejestrze znalazły się tylko te żyjące w momencie powstania PESEL-u. Ponadto, niezależnie od kompletności danych w systemie PESEL, wiadomo, że w kolejnych latach liczba nadanych imion jest zależna od czynników biologicznych i demograficznych – każdego roku przecież rodzi się różna liczba osób, a dodatkowe zróżnicowanie wynika z faktu, że inna jest liczba osób odmiennych płci (w Polsce zwykle rodzi się więcej kobiet niż mężczyzn). Podobnie rzecz wygląda w przypadku badania zróżnicowania geograficznego imiennictwa pod kątem wskaźników ilościowych – w poszczególnych regionach (które możemy definiować np. jako województwa, powiaty, miasta itd.) również mamy do czynienia z różną liczbą osób. Na rys. 17 i 18 przedstawiono liczebność żeńskich i męskich korpusów imienniczych w poszczególnych latach z przedziału 1901–2010.

Jak widać, wielkość korpusu imienniczego dla poszczególnych lat może różnić się diametralnie. Przykładowo dla 1901 r. korpus imion męskich liczy 16 489, imion żeńskich – 37 942, natomiast w 1983 r. jest to odpowiednio 364 031 i 347 812.

Rys. 17. Liczebność korpusów imion żeńskich w poszczególnych latach z przedziału 1901–2010

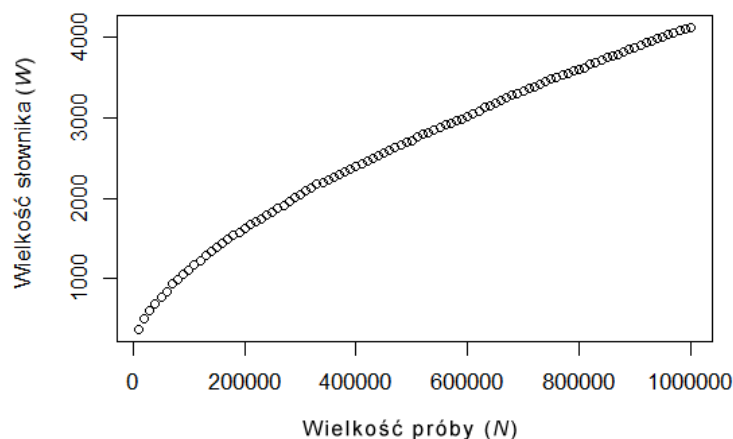


Rys. 18. Liczebność korpusów imion męskich w poszczególnych latach z przedziału 1901–2010



W badaniach z zakresu statystyki leksykalnej o charakterze porównawczym, w których do zbadania różnic lub podobieństw między tekstami wykorzystuje się wskaźniki ilościowe, powszechnie znany jest problem długości porównywanych tekstów. Zwraca się uwagę na konieczność operowania tekstami o możliwie równej długości, nie istnieje bowiem żadna miara bogactwa leksykalnego tekstu, która byłaby niezależna od jego długości (Tweedie, Baayen 1998: 323–324, Baayen 2001: 24 i n., Pawłowski 2003: 175, Rybicki 2008: 184) – wszystkie w jakimś stopniu opierają się na wzajemnej relacji długości tekstu (N) i wielkości jego słownika (W), określanej mianem *type-token ratio* (TTR). Jak wiadomo, wielkość słownika, czyli liczba słów użytych przez autora w tekście, którą można traktować jako podstawowy wskaźnik ilościowego bogactwa tekstu, jest funkcją rosnącą długości tekstu (Sambor 1972: 219, Hout, Vermeer 2007: 123), choć w miarę zwiększania się długości tekstu

Rys. 19. Zależność wielkości słownika od wielkości próby



przyrosty słownictwa są coraz mniejsze. Nie inaczej jest w przypadku korpusów imion nadanych w określonym czasie czy miejscu. Tu również wraz ze wzrostem liczby osób, którym nadano imiona, zwiększa się wielkość słownika imion, których użyto do nominacji (Huschka, Wagner 2010: 17). Obrazuje to wykres na rys. 19, wygenerowany na podstawie losowej próby miliona osób z naszego zbioru. Punkty na wykresie odpowiadają skumulowanej wielkości słownika kolejnych 10-tysięcznych fragmentów korpusu. Wziąwszy pod uwagę powyższą właściwość, porównując tak znacząco różniące się liczbowo korpusy imion, jak np. z lat 1901 i 1983 (zob. rys. 17 i 18), siłą rzeczy wartości np. wielkości słownika będą zawsze wyższe w przypadku dużo większego korpusu.

Dysponując tekstami lub, w naszym przypadku, zbiorami imion o różnej długości, możemy zastosować dwa rozwiązania problemu nierównych prób – 1) wyrównać je pod względem wielkości lub też 2) zastosować bardziej zaawansowane modele statystyczne polegające na korekcie szacowanych miar przy wykorzystaniu skomplikowanych matematycznych metod (Baayen 2008: 224–226, por. też Jarvis 2002), co jednak Huschka i Wagner (2010: 18) trafnie określają jako „naukę samą w sobie”. Dążenie do zastosowania drugiej z metod w badaniach nad tekstami języka naturalnego wynika z trudności zredukowania ich do mniejszej podpróby – tekst ma określoną strukturę kompozycyjną, którą trudno odtworzyć na podstawie losowania, nie mówiąc już o „obcięciu” którejś z jego części. W przypadku pojedynczych imion zaczerpniętych z określonych rejestrów stworzenie mniejszych losowych podprób jest mniej problematyczne. Oczywiście rozwiązanie takie niesie ze sobą konieczność zrezygnowania z części danych (część imion, zwłaszcza bardzo rzadkich z pewnością zostanie pominięta), jednak pozwala na porównanie struktury liczbowej zbiorów imienniczych w takich samych warunkach i wydaje się to być dopuszczalnym ograniczeniem. O metodach tworzenia równych podprób zastosowanych w niniejszej pracy piszę w podrozdziale III.2.

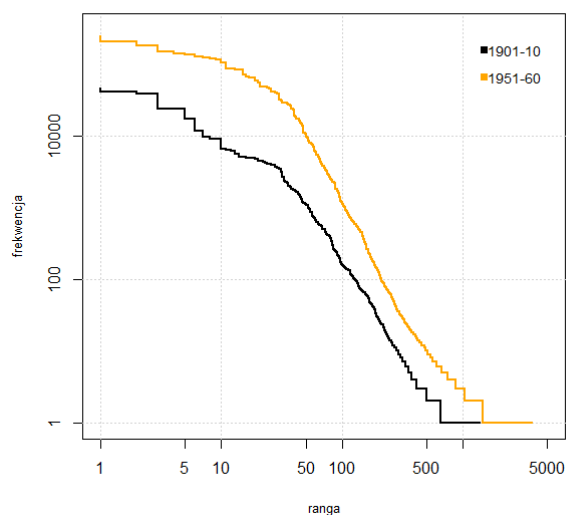
W celu zilustrowania, w jaki sposób różnica w wielkości próby może wpływać na ewentualne wnioski na temat różnic lub podobieństw między dwoma zbiorami antroponimów w różnych okresach lub miejscach, przedstawię obliczenia kilku wybranych podstawowych charakterystyk ilościowych dla różniących się znacznie pod względem wielkości zbiorów imion męskich z dwóch dekadach: 1901–10 (ok. 400 tys. jednostek) i 1951–1960 (ok. 3,4 mln jednostek) zestawionych z obliczeniami tych samych charakterystyk po redukcji większego ze zbiorów (tab. 1).

Tab. 1. Wybrane wskaźniki dla imion męskich w dwóch dekadach (przed i po redukcji większej próby)

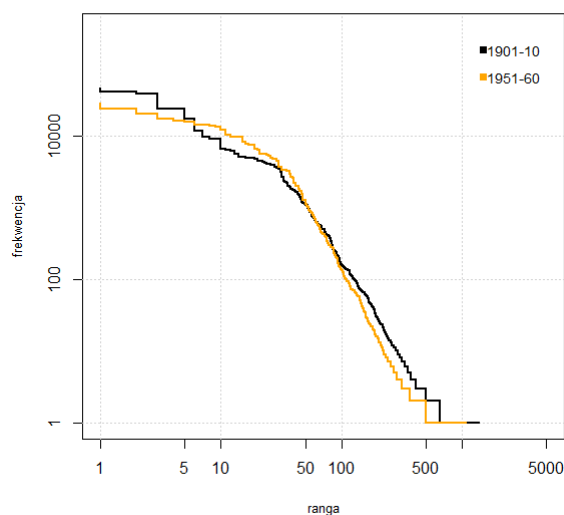
	1901–10	1951–60	1951–60 po redukcji
wielkość zbioru	393 874	3 402 721	393 874
liczba typów	1366	3724	1067
średnia liczba użytkowników	285	910	369
wskaźnik koncentracji Lorenza	0.969	0.988	0.961

Jak łatwo zauważyć, pozornie ogromne różnice między danymi przed redukcją większej próby, stają się dużo mniej istotne (choć nadal występują), gdy porównujemy próby o równej wielkości. Różnice te wyraźnie odbijają się także na graficznym obrazie rozkładu częstości, jaki stanowi wykres Zipfowskiej zależności między rangą a częstością wyrazów (imion) (rys. 20 i 21). Na rys. 20 wykresy dla poszczególnych dekad mają podobny kształt, jednak znacznie od siebie odbiegają, co znacznie utrudnia ich porównanie. Dopiero na rys. 21 możliwe jest prześledzenie, jak „zachowują się” imiona w poszczególnych klasach częstości w tych samych warunkach, a więc bez wpływu wielkości próby na szereg zjawisk od niej zależnych.

Rys. 20. Rozkład częstości imion męskich dla dwóch dekad przed redukcją większej próby



Rys. 21. Rozkład częstości imion męskich dla dwóch dekad po redukcji większej próby

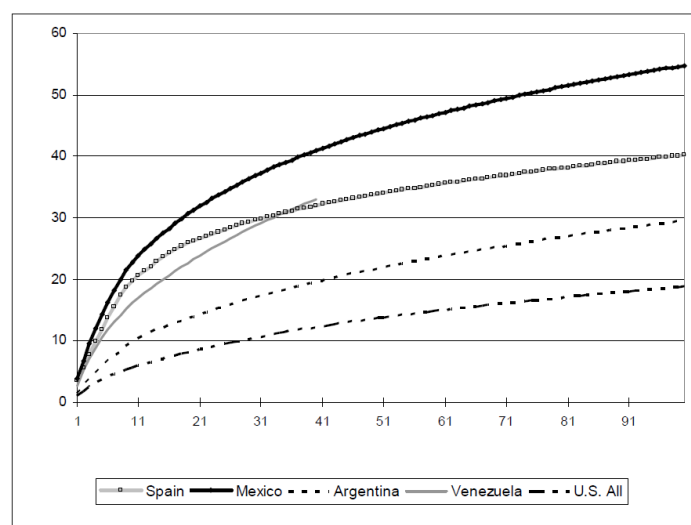


Na problem nierówności prób w badaniach antroponimicznych zwrócili uwagę niemieccy badacze D. Huschka, J. Gerhards i G. Wagner (2009) w pracy na temat różnic w imiennictwie Niemiec Wschodnich i Zachodnich. Przyjęli oni hipotezę o bardziej zindywidualizowanym charakterze imiennictwa w RFN, bazując na przekonaniu, 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 przeczyło przyjętej hipotezie i wręcz dowodziło czegoś odwrotnego. Po zredukowaniu większego korpusu dla RFN do wielkości korpusu 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 wystąpiło np. więcej imion słowiańskich), pod względem wielkości słownika imion niewiele się one różniły (Huschka i in. 2009: 213–214).

W wielu dotychczasowych pracach antroponimicznych, w których zajmowano się kwestiami ilościowej charakterystyki zbiorów antroponimicznych, niestety, kwestia równości porównywanych prób rzadko była brana pod uwagę, a wnioski w nich zawarte wymagają ponownej weryfikacji. Przykładem może być praca P. Mateosa i K. Tuckera (2008), w której autorzy dokonali porównania danych dla 100 najczęstszych nazwisk w Hiszpanii z analogicznymi danymi dla innych krajów hiszpańskojęzycznych oraz dla Stanów Zjednoczonych. Dane pochodziły z różnych rejestrów (książki telefoniczne, spisy wyborców, spisy powszechne). Celem badania było określenie w jakim stopniu antroponimy o rangach od 1 do 100 pokrywają sumę użyć imion w poszczególnych państwach. Po naniesieniu danych na wykres (rys. 22) okazało się, że występują między nimi ogromne rozbieżności. Autorzy uznali, że dane dla USA i Argentyny są bardziej spójne niż w przypadku pozostałych trzech państw, a fakt ten powiązali z większym napływem imigrantów (a więc i obcych nazwisk) do tych krajów (Mateos, Tucker 2008: 179–181). Oczywiście nie można wykluczać

Rys. 22. Kumulacyjny procentowy udział 100 najczęstszych nazwisk w wybranych krajach (za: Mateos, Tucker 2008: 181)



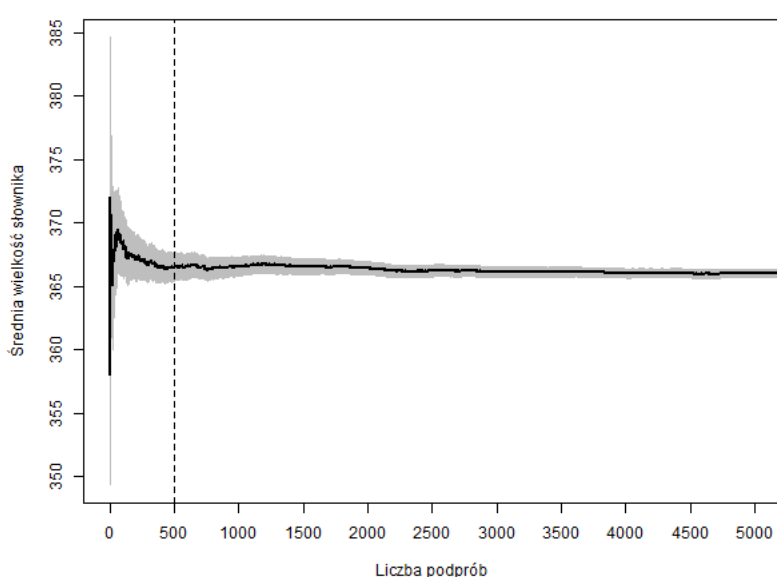
wpływu czynników migracyjnych, o których piszą autorzy, jednak należy zachować dużą ostrożność w przyjmowaniu takiego wniosku, gdyż badane próby różniły się jakością i ilością danych (nie wiadomo w jakim stopniu, gdyż autorzy nie podają takich informacji).

Wspomniana praca powieliła błąd z wcześniejszych tekstów Tuckera (2001, 2002, 2004) o imionach i nazwiskach w Kanadzie i Stanach Zjednoczonych, na co uwagę zwrócili już Huschka i Wagner (2010: 20). Również w kilku polskich pracach na temat imiennictwa dokonywano obliczeń określonych charakterystyk liczbowych na nierównych próbach, co również mogło skutkować wyciągnięciem błędnych wniosków. W pracy na temat zmian frekwencyjnych w zasobie imion w Polsce do 1994 r. K. Nowik (1998), porównując ponad 19-milionowy zbiór imion z lat 1901–94 z niespełna 3-milionowym zbiorem antroponimów z lat 1981–90, zauważyła, że 20 pierwszych imion w pierwszym zbiorze pokrywało ponad 51% sumy użyć imion, a w drugim zbiorze z lat 1981–90 było to już ok. 66%. Na tej podstawie badaczka stwierdziła, że w latach 80. przy wyborze imienia w większym stopniu kierowano się „modą”, niż miało to miejsce na przestrzeni całego XX w. (Nowik 1998: 62). Błąd ten powielił P. Swoboda w artykule na temat frekwencji i geografii najczęstszych imion w Polsce na przestrzeni lat 1995–2010, zestawiając dane z tego okresu (obejmujące po ok. 3 mln okazów) z danymi z pracy K. Nowik (Swoboda 2013: 23–32). Również S. Gajda w przywoływanym już wielokrotnie artykule, porównywał pod kątem bogactwa leksykalnego zbiory różnej wielkości (tj. o sumie użyć np. 1016, 1853, 2319), zaznaczając co prawda, że liczba osób w danej grupie wpływa na liczbę użytych imion, jednak za wystarczające do rozwiązania tego problemu uznał wyliczenie średniej liczby użytkowników jednego imienia (Gajda 1973: 269), co – w świetle przytoczonych informacji o zależności wszystkich wskaźników od wielkości korpusu – nie wytrzymuje krytyki.

III.2. Skalowanie zbiorów imion, wielkość i liczba podprób

Wspominani wcześniej D. Huschka i G. Wagner zaproponowali metodę skalowania korpusów imion (Huschka, Wagner 2010: 19–20), która została zastosowana w niniejszej pracy. Chcąc porównywać ze sobą dwa lub więcej korpusów imion (np. nadawanych w poszczególnych przedziałach czasowych lub na różnych obszarach geograficznych), należy wyrównać je pod względem długości, czyli zredukować je do wielkości najmniejszego z nich (lub mniejszej). W tym celu należy sporządzić z każdego z nich podpróby równej wielkości metodą losowania bez zwracania. Przy czym trzeba mieć na uwadze, że nie wystarczy wylosowanie jednej podpróby z jednego korpusu. Obliczając np. wielkość słownika danej podpróby, musimy mieć świadomość, że wielkość słownika w kolejnej podpróbie wylosowanej z tego samego korpusu może się znacznie różnić. Dlatego procedurę tę należy powtórzyć odpowiednią ilość razy, aż do momentu, gdy średnia dla wszystkich podprób dla danego korpusu się ustabilizuje. W celu określenia, jaka liczba podprób w naszym przypadku jest wystarczająca, sporządziłem 10 000 podprób (każda licząca po 40 tys. okazów⁴⁹) dla największego zbioru imion, tj. dla imion męskich nadanych w 1983 r. (liczącego pierwotnie ponad 360 tys. okazów). Dla każdej z podprób obliczono liczbę różnych imion w niej występujących (wielkość słownika). Zaledwie pierwszych dziesięć wyników (358, 376, 380, 374, 358, 368, 386, 368, 349, 363) pokazuje, że wielkość słownika podlega pewnym wahaniom między wylosowanymi podpróbami, a więc dla konkretnego roku musimy

Rys. 23. Średnia wartość wielkości słownika (os y) obliczona na podstawie zwiększającej się liczby podprób (os x).



⁴⁹ Uzasadnienie wyboru takiej wielkości podpróby znajduje się w dalszej części niniejszego podrozdziału.

wyliczyć średnią z odpowiedniej liczby podprób. Na wykresie na rys. 23 czarna linia obrazuje średnią wielkość słownika (oś y) obliczoną na podstawie kolejnych losowanych podprób (oś x), szary obszar powyżej i poniżej tej krzywej odpowiada 95% przedziałom ufności. Widzimy, że wartości średniej dla pierwszych kilkudziesięciu podprób podlegają dużym wahaniom i dopiero po kilkuset losowaniach zaczynają się stabilizować. W pracy przyjąłem, że wskaźniki bogactwa leksykalnego dla poszczególnych lat będę obliczał na podstawie 500 losowych podprób, a zatem przy porównywaniu imiennictwa poszczególnych okresów lub regionów będę operował średnimi wartościami dla danego wskaźnika.

Osobną kwestią jest optymalna wielkość podpróby. W badaniach z zakresu stylometrii często zadaje się pytanie o minimalny rozmiar próby tekstu danego autora, który pozwoli na otrzymanie miarodajnych wyników np. atrybucji autorskiej (Ruszkowski 2003, Eder 2010, Luyckx, Daelemans 2011). Wydaje się oczywiste, że im większy (pełniejszy korpus), tym wyniki obliczeń są bardziej miarodajne (López-Escobedo i in. 2013: 608), jednak z uwagi na fragmentaryczność danych, którymi dysponujemy (zwłaszcza dla wcześniejszych lat), musimy sporządzić odpowiednio mniejsze próby. Jak już wspomniano, w przypadku imiennictwa ogólnopolskiego najmniejszą liczbą danych dysponujemy w przypadku imion męskich z 1901 r. (nieco ponad 16 tys.). W związku z tym początkowo zamierzałem przeprowadzić obliczenia dotyczące imiennictwa poszczególnych lat w skali ogólnopolskiej w oparciu o podpróby wielkości 15 tys. Postanowiłem jednak sprawdzić najpierw, w jakim stopniu zmienność wartości wskaźników w kolejnych latach otrzymanych dla tak małych podprób odzwierciedla zmienność w przypadku stosunkowo dużych, 100-tysięcznych podprób. W tym celu obliczono współczynnik korelacji liniowej Pearsona⁵⁰ między średnimi wartościami kilku wybranych podstawowych wskaźników bogactwa leksykalnego w kolejnych latach z przedziału 1920–2010⁵¹, otrzymanych na podstawie dużych – 100-tysięcznych podprób, a średnimi wartościami wyliczonymi na podstawie 15-, 20- i 40-tysięcznych podprób. Wyniki obliczeń zamieszczono w tabeli 2.

Choć we wszystkich przypadkach mamy do czynienia z silną korelacją ($\rho > 0,9$), to jednak dopiero w przypadku ostatniej kolumny możemy mówić o niemal pełnej korelacji, a zatem

⁵⁰ Współczynnik korelacji Pearsona (ρ) mierzy liniową zależność między dwoma zestawami danych. Może przyjmować wartości z przedziału $[-1, 1]$. Im większa wartość bezwzględna ρ , tym większa jest korelacja między dwoma zmiennymi. Przykładowo wartość $\rho = 0$ oznacza brak zależności, $\rho = 1$ oznacza dokładną (pełną) korelację, zaś $\rho = -1$ oznacza pełną korelację ujemną (zależność między zmiennymi jest wówczas odwrotnie proporcjonalna).

⁵¹ Wynika to z faktu, iż dopiero dla 1920 r. dysponujemy w przypadku imion męskich korpusem o wielkości większej niż 100 tys. (patrz ryc. 18).

układ wartości wskaźników obliczonych na podstawie 40-tys. próbek w kolejnych latach najlepiej odzwierciedla układ tych wartości dla dużej, stutysięcznej próby. Postanowiłem zatem, by porównań między poszczególnymi latami dokonywać przy pomocy średnich otrzymanych dla próbek o $N = 40000$, co jednak spowodowało konieczność rezygnacji z analizy wskaźników dla lat 1901–1905, gdyż dopiero dla roku 1906 dysponujemy korpusem imion męskich o takiej wielkości (dokładnie 40 465).

Tab. 2. Współczynnik korelacji liniowej między zbiorami różnej wielkości a zbiorem stutysięcznym w zakresie realizacji wybranych wskaźników bogactwa leksykalnego w kolejnych latach z przedziału 1920–2010

	$N = 15000$	$N = 20000$	$N = 40000$
wielkość słownika	0,9787	0,9841	0,995
wsk. oryginalności Guirauda	0,9787	0,9841	0,995
wsk. koncentracji Lorenza	0,9335	0,9475	0,9818

Badanie zróżnicowania geograficznego w zakresie wskaźników bogactwa leksykalnego na obszarze poszczególnych województw musiałem niestety ograniczyć tylko do lat 1951–2010, co wynikało z trudności rozpatrywania materiału imienniczego sprzed II wojny światowej w kontekście powojennych i współczesnych podziałów administracyjnych. Ponadto materiał imienniczy nie jest tu analizowany w skali rocznej, a w następujących po sobie dekadach (1951–1960, 1961–1970, 1971–1980, 1981–1990, 1991–2000, 2001–2010), czego powodem była dosyć niska liczebność rocznych korpusów w niektórych województwach. Udało się jednak dzięki temu uzyskać równe, 10-letnie przedziały, a liczebność korpusów, tak jak w wypadku badania imiennictwa ogólnopolskiego, wyniosła 40 tys.

W takich samych przedziałach czasowych, jednak nieco mniejsze próbki sporządzone zostały z kolei w celu porównania liczbowych wskaźników bogactwa leksykalnego dla wybranych największych miast Polski (Krakowa, Poznania, Warszawy i Wrocławia) na tle województw, w których się znajdują. Tu zmuszony byłem do operowania 30-tysięcznymi próbkami, gdyż w przypadku niektórych dekad, liczebność oryginalnych korpusów w poszczególnych miastach (poza Warszawą) nie wyniosła 40 tysięcy.⁵² Jak już wspomniano wcześniej, obliczenia dla województw zostały dokonane z wyłączeniem materiału z obszaru wymienionych miast.

⁵² Z tego względu trzeba było zrezygnować np. badania porównawczego dla Łodzi i woj. łódzkiego, gdyż w tym mieście wielkość korpusu w większości dekad nie przekraczała nawet 20 tys. okazów.

III.3. Wyniki analizy parametrów rozkładu częstości

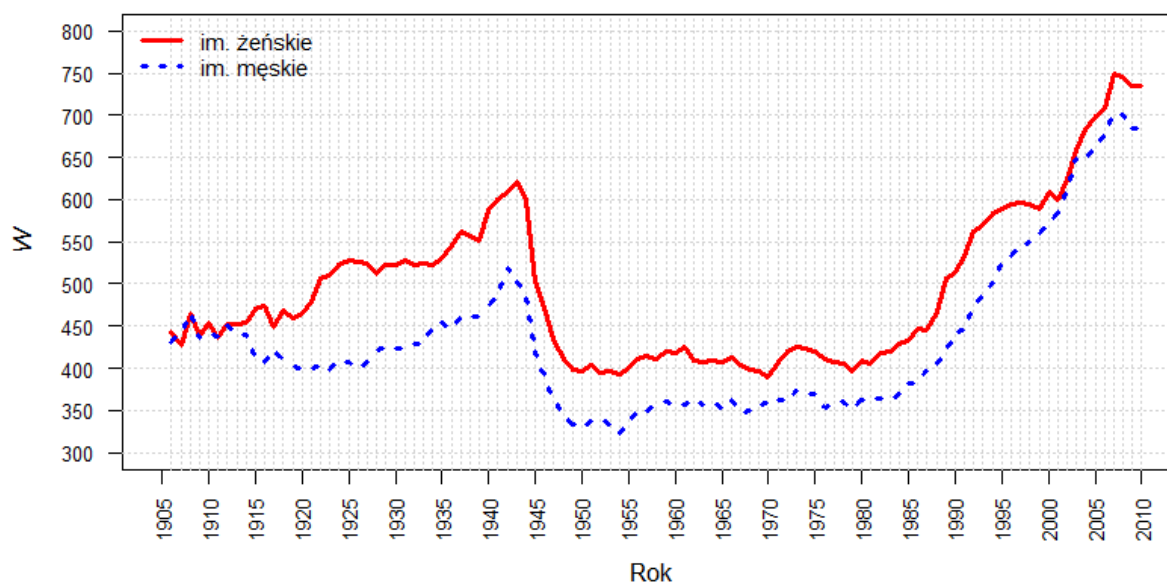
III.3.1. Wielkość słownika

Wśród wskaźników bogactwa leksykalnego za najbardziej podstawowy uważa się wielkość słownika tekstu (W). Jest to liczba różnych słów (typów), które w danym tekście wystąpiły (Tweedie, Baayen 1998: 325)⁵³. Interpretacja tego wskaźnika jest oczywista – im większy słownik danego tekstu, tym bogatsze, bardziej różnorodne jest jego słownictwo. Adaptując tę miarę do badania ilościowego bogactwa imiennictwa, wielkością słownika będzie liczba różnych imion, których użyto do nominacji osób ujętych w określonym zbiorze.

A) Wielkość słownika w imiennictwie ogólnopolskim (lata 1906–2010)

Dane o liczbie typów imion nadawanych kobietom i mężczyznom w całej Polsce w poszczególnych latach z przedziału 1906–2010 zostały naniesione na wykres na rys. 24.

Rys. 24. Wielkość słownika w imiennictwie ogólnopolskim w kolejnych latach

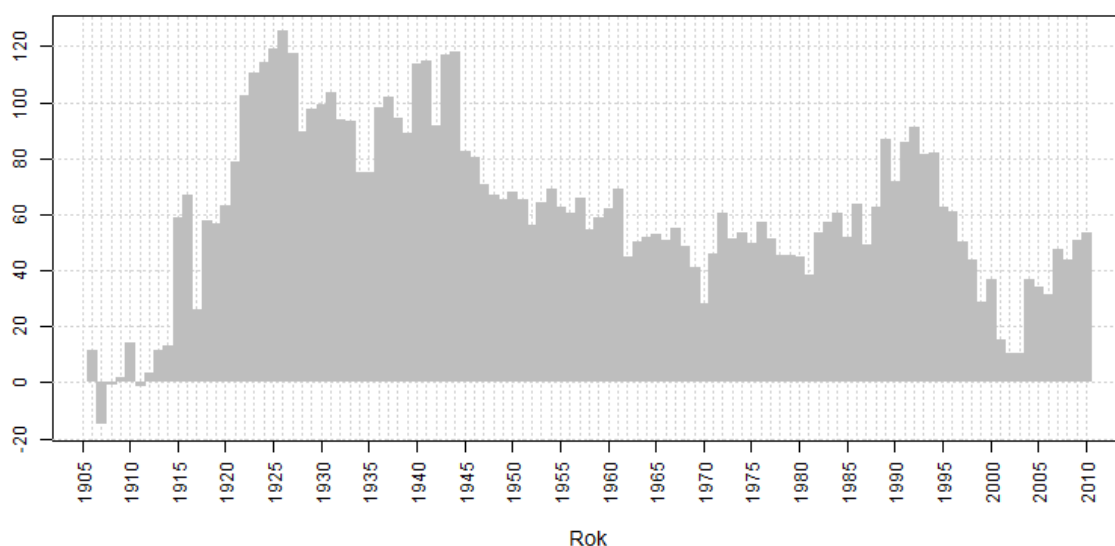


Krzywe na powyższym wykresie ilustrują zmienność liczby typów (czyli wielkości słownika) w poszczególnych latach. Kierunek przebiegu obu krzywych jest podobny (współczynnik korelacji Pearsona dla obu grup wyniósł 0,947). W obu przypadkach mamy do

⁵³ W badaniach z zakresu statystyki leksykalnej w przypadku porównywania tekstów o zbliżonej, jednak nie równej długości, oblicza się na ogół średnią częstość wyrazu w tekście ($\bar{f} = N/W$), określaną również mianem indeksem powtarzalności (Kamińska-Szmaj 1990: 12) lub wskaźnikiem iteracji (Kowalska 2013: 57). W przypadku niniejszego badania, w którym porównuję podpróby równej wielkości (40 tys.), wystarczające jest operowanie liczbą typów, czyli wielkością słownika.

czynienia ze stopniowym wzrostem liczby typów od lat 1918–1920 aż do lat 1942–1943, gdy osiągnęła ona poziom 518 (imiona męskie) i 620 (imiona żeńskie). Następnie obserwujemy znaczący spadek omawianego wskaźnika, który w okresie od końca lat 40. do końca lat 70. utrzymywał się na stosunkowo stałym i niskim poziomie 325–375 dla imion męskich oraz 400–450 dla imion żeńskich. Od początku lat 80. widoczny jest stały trend wzrastający w zakresie wielkości słownika, który w latach dwutysięcznych przekroczył maksymalny poziom z poprzedniego cyklu wzrostowego z lat 40., by pod koniec badanego okresu, w 2007 r., osiągnąć wartość 702 (im. męskie) i 750 (im. żeńskie).

Rys. 25. Różnica w wielkości słownika między imiennictwem żeńskim i męskim (wartości dodatnie oznaczają większy słownik imion żeńskich)



O ile przebieg zmian w zakresie liczby typów miał podobny charakter w obu grupach, to łatwo zauważyć, że wartości tego wskaźnika dla imion męskich i żeńskich wyraźnie od siebie odbiegają. Na rys. 25 zamieszczono wykres słupkowy obrazujący poziom różnic między wielkością słownika imion żeńskich i męskich ($W_z - W_m$). Z wyjątkiem początkowych lat (1906-1911) przez resztę rozpatrywanego okresu imiennictwo żeńskie zawsze charakteryzowało się większą liczbą typów imion niż imiennictwo męskie, choć oczywiście różnie kształtował się poziom zróżnicowania między obiema grupami. Początkowo był on niewielki (wahając się między 1 a 15 imionami), od roku 1915 widoczny jest znaczny wzrost, natomiast od lat 20. do początku lat 40. notowane były największe różnice – maksymalna liczba różnic wystąpiła w 1926 r., gdy odnotowano ponad 125 więcej typów imion żeńskich (525,8) niż męskich (400,3). W kolejnych latach zróżnicowanie stopniowo malało do poziomu między 40–60 typów i dopiero na przełomie lat 80. i 90. wzrosło do ok. 80-90 typów. Po tym okresie mamy do czynienia z kolejnym spadkiem zróżnicowania do poziomu

między 15 a 50 typów. Możemy zatem stwierdzić, że na przestrzeni ostatnich 70 lat doszło do stopniowej niwelacji różnic pomiędzy liczbą typów w imiennictwie męskim i żeńskim, niemniej nadal w zakresie tego wskaźnika obserwujemy większą różnorodność w przypadku imion żeńskich. Warto przytoczyć w tym miejscu fakt, iż zjawisko przewagi imiennictwa żeńskiego nad męskim w zakresie wielkości słownika nie dotyczy jedynie Polski i wydaje się mieć szerszy zasięg. Odnotowali je także autorzy prac na temat XX-wiecznego imiennictwa w Niemczech (Huschka, Wagner 2010: 28), Norwegii (Alhaug 2002: 311) Stanach Zjednoczonych (Hahn, Bentley 2003: 122) czy Wielkiej Brytanii (Tucker 2004: 39), choć należy pamiętać, że ich badania dotyczyły jedynie wybranych punktów czasowych lub krótszych okresów, nie wiemy zatem, czy omawiane zjawisko ma tam charakter długotrwały.

B) Wielkość słownika w imiennictwie na obszarze współczesnych województw (lata 1951–2010)

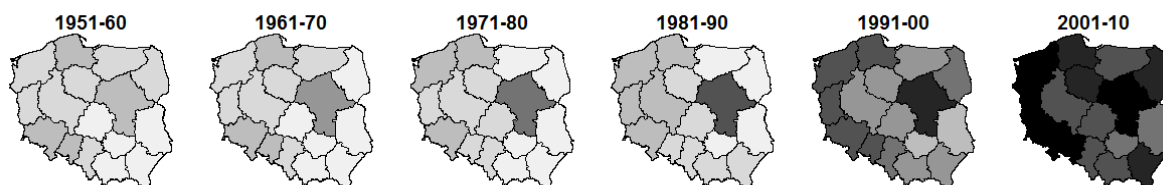
Dane zamieszczone w tabeli 3 i naniesione na mapę administracyjną (rys. 26) już przy pobieżnej analizie przejawiają kilka generalnych tendencji. Łatwo zaobserwować, że w pierwszych czterech dekadach (a zwłaszcza w latach 1961–1990) znacznie większa liczba typów (zarówno męskich i jak i żeńskich) występuje w zachodniej części Polski niż na jej wschodzie, na którego tle wyraźnie odcina się jednak województwo mazowieckie. Zwarty obszar, który cechuje się najmniejszymi słownikami imion stanowią zwłaszcza województwa południowo-wschodnie (małopolskie, podkarpackie, lubelskie i świętokrzyskie). Z kolei najwięcej typów notowano przede wszystkim na obszarze zewnętrznego pasa województw zachodnich (od śląskiego na południu po pomorskie na północy). W kolejnych dwóch dekadach przewaga w zakresie wielkości słownika dalej widoczna jest w przypadku województwa mazowieckiego i województw zachodnich, jednak podział na wschód i zachód nie jest już tak ostry jak wcześniej, co pokazuje znaczny wzrost słownika na obszarze województw podkarpackiego i podlaskiego. Fakt zdecydowanego odchylenia województwa mazowieckiego na tle sąsiadujących obszarów wiąże się zapewne z wysoką liczbą typów w największym z jego miast, czyli w Warszawie, co postaram się zweryfikować w dalszej części, poświęconej analizie badanego wskaźnika w największych miastach na tle województw.

Tab. 3. Średnia wielkość słownika na obszarze współczesnych województw

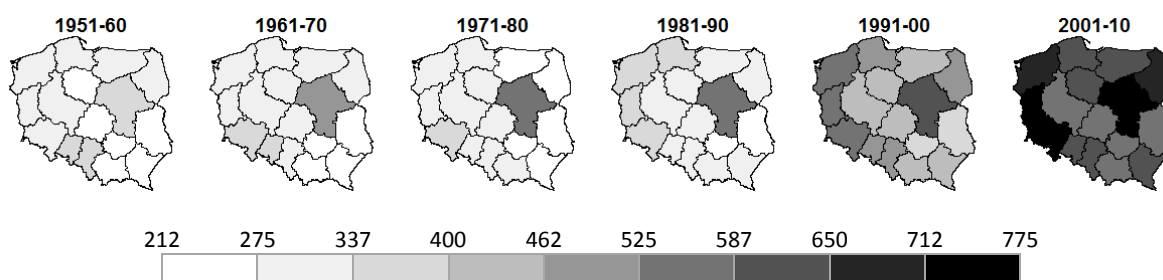
woj.	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	405	400	410	442	632	775	322	342	339	371	559	713
KUJ.	340	345	350	371	480	656	270	279	281	301	446	619
LBL.	295	286	299	312	431	581	242	250	240	262	384	546
LBS.	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.	289	312	336	359	483	602	234	259	284	309	446	584
MAZ.	426	476	539	621	669	749	394	467	553	586	610	720
OPL.	427	408	383	397	639	728	370	358	325	323	520	649
PKD.	286	302	306	350	514	665	212	237	262	282	454	632
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
ŚLS.	386	390	367	393	531	636	342	337	315	324	471	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	13,64	14,11	17,54	18,84	13,31	9,93	17,22	18,73	23,65	22,62	13,29	9,85

Rys. 26. Średnia wielkość słownika (dane z tab. 3)

imiona żeńskie



imiona męskie



Podobny podział na osi wschód-zachód (choć nieco mniej wyraźny i dotyczący głównie czterech środkowych dekad) możemy też zaobserwować jeśli chodzi o różnicę między wielkością słownika imion męskich i żeńskich. Choć podobnie jak w imiennictwie ogólnopolskim niemal zawsze imiennictwo żeńskie odznaczało się większą liczbą typów, to jednak w województwach wschodnich i centralnych (mazowieckim, małopolskim, podlaskim, lubelskim i świętokrzyskim) zwykle różnice w tym zakresie były dużo mniejsze

niż w województwach zachodnich. W czterech pierwszych dekadach najmniej różnic notowano właśnie w woj. mazowieckim. Co więcej – tylko tam doszło do jedyne go przypadku, gdy imiennictwo męskie było bogatsze od żeńskiego (w dekadzie 1971–80). W przypadku większości województw możemy mówić raczej o zmniejszaniu się różnic między wielkością słowników imion męskich i żeńskich na przestrzeni całego badanego okresu (w największym stopniu dotyczy to województw wschodnich, ale w ostatniej dekadzie również województwa lubuskiego), natomiast w woj. łódzkim i opolskim tendencja ta przybrała kierunek odwrotny.

Analizując wartości wskaźnika zmienności v , możemy stwierdzić, że tak w obrębie imiennictwa żeńskiego, jak i męskiego, poziom regionalnego zróżnicowania w zakresie wielkości słownika do lat 80. wzrastał, przy czym należy zaznaczyć, że większe różnice wystąpiły w przypadku imiennictwa męskiego. Natomiast w ostatnich dwóch dekadach różnice te, choć nadal widoczne, uległy znacznemu zmniejszeniu – w ostatniej dekadzie odnotowano najniższą wartość wskaźnika v w stosunku do okresów poprzedzających. Widzimy również, że inaczej niż wcześniej, imiennictwo męskie i żeńskie odznacza się już niemal takim samym stopniem zróżnicowania (możemy nawet mówić o nieznacznie większych wartościach wskaźnika v dla imiennictwa żeńskiego).

C) Wielkość słownika w imiennictwie wybranych miast na tle województw (lata 1951–2010)

Tab. 4. Średnia wielkość słownika w ośrodkach wielkomiejskich na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	382	396	405	455	599	669	324	378	352	410	544	639
dśl.	344	344	348	370	530	667	274	280	280	297	463	601
różnica	38	52	57	85	69	2	50	98	72	113	81	38
Kraków	312	334	335	389	462	530	276	293	304	343	422	478
młp.	247	262	282	287	405	512	194	213	230	246	372	516
różnica	65	72	53	102	57	18	82	80	74	97	50	-38
Warszawa	547	649	704	813	831	795	524	657	755	778	754	766
maz.	268	264	265	273	346	448	228	227	222	231	308	425
różnica	279	385	439	540	485	347	296	430	533	547	446	341
Poznań	350	372	365	433	506	542	303	302	328	356	470	530
włp.	293	311	318	330	440	538	246	244	243	271	362	489
różnica	57	61	47	103	66	4	57	58	85	85	108	41

Z danych zamieszczonych w tab. 4, dotyczących różnic pomiędzy liczbą imion nadawanych w największych miastach na tle województw, w których się znajdują, wynika, że

niemal zawsze miasta odznaczały się większym słownikiem imion. Zdecydowanie największą różnicę obserwujemy w przypadku Warszawy – mowa tu o słowniku czasem kilkukrotnie większym niż na obszarze pozostałej części woj. mazowieckiego. Potwierdza to zatem przypuszczenia, że odnotowane wcześniej znaczne odchylenia we wskazaniach dla tego województwa na tle sąsiadujących województw są efektem wpływu wielkości słownika imion samego miasta stołecznego. W pozostałych badanych miastach różnice te nie są aż tak uderzające – największą odnotowano dla imion męskich w przypadku Krakowa i woj. małopolskiego w okresie 1951–60. Najmniejsze różnice występowały między imiennictwem Wrocławia i woj. dolnośląskiego. W ostatniej dekadzie różnice między miastami a prowincją jest już minimalna (nie dotyczy to oczywiście Warszawy, gdzie nadal utrzymuje się spora różnica), nieprzekraczająca 9%. Co więcej – w Małopolsce doszło do odwrócenia sytuacji – słownik imion nadawanych na obszarze województwa był o ok. 7% większy niż w Krakowie. Podobnie jak to miało miejsce w przypadku wskaźnika zmienności dla województw, również i tu obserwujemy większe różnice między miastem i województwem z reguły w obrębie imiennictwa męskiego.

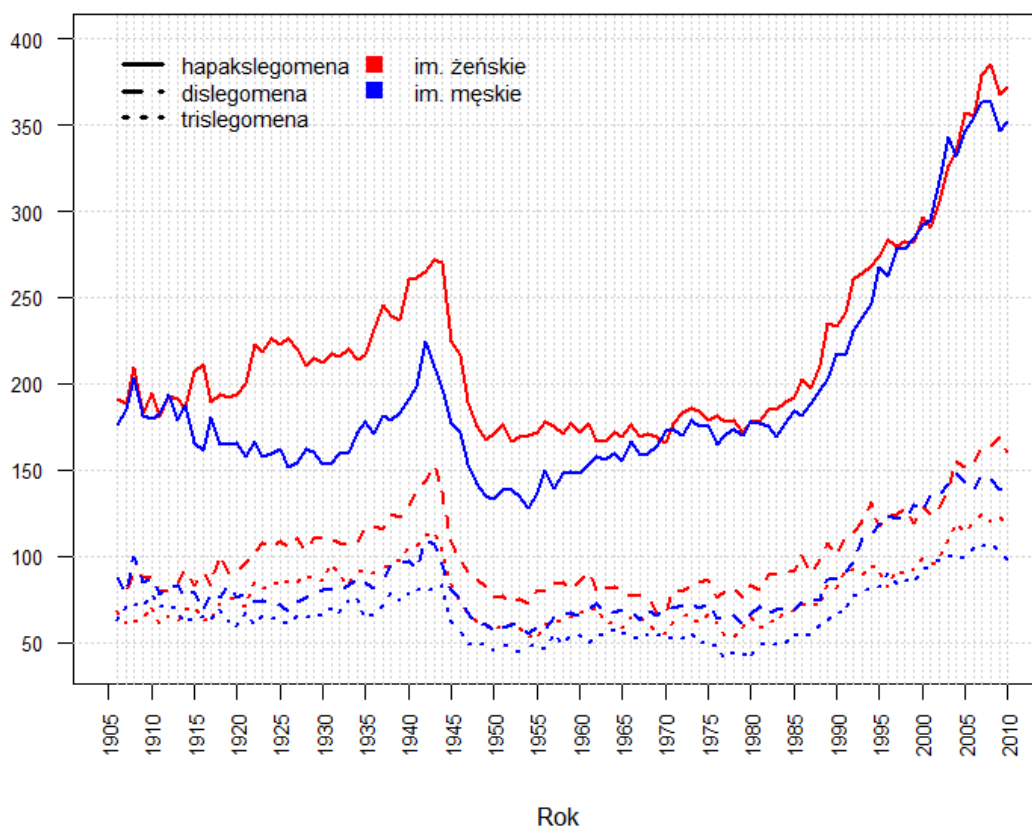
Jeśli chodzi o różnice w wielkości słownika między imiennictwem żeńskim i męskim, to w Krakowie, Wrocławiu i Poznaniu i otaczających je obszarach, a także w województwie mazowieckim zdecydowanie więcej typów zwykle (poza ostatnią dekadą w woj. małopolskim) notowano w obrębie imion żeńskich, natomiast w Warszawie w latach 1961–80 mieliśmy do czynienia z sytuacją odwrotną. W Poznaniu (poza czwartą dekadą) i Wrocławiu różnice te zawsze były jednak zawsze mniejsze niż na pozostałym obszarze województw, których są stolicami. W Krakowie i Warszawie taka tendencja utrzymywała się przez trzy pierwsze dekady, by następnie ulec odwróceniu.

III.3.2. Liczba hapaks-, dis- i trislegomenów

W badaniach stylometrycznych poza wielkością słownika jednym z najprostszych wskaźników bogactwa leksykalnego tekstu jest udział w nim leksyki o najniższej frekwencji (por. np. Hoover 2003: 152, Malvern i in. 2004: 125, Moskovich 2016: 122–125). Do najrzadszej leksyki z reguły zalicza się tzw. hapaks-, dis- i trislegomena⁵⁴, czyli wyrazy o częstościach odpowiednio: $f = 1, f = 2, f = 3$.

A) Liczba hapaks-, dis- i trislegomenów w imiennictwie ogólnopolskim (lata 1906–2010)

Rys. 27. Liczba hapaks- dis- i trislegomenów w imiennictwie ogólnopolskim



Jak łatwo zauważyć, krzywe na wykresie ilustrującym liczbę hapaks-, dis- i trislegomenów (rys. 27) w zbiorach imion z kolejnych lat, mają podobny wygląd do wykresu dla liczby typów (por. rys. 24), zwłaszcza w przypadku liczby hapakslegomenów, co świadczy o wyraźnym sprzężeniu tych wskaźników. Wynika to z faktu, że o wzroście wielkości słownika decydują właśnie nowe, rzadkie jednostki dodane do niego, a nie imiona

⁵⁴ K. Skowronek (2001: 83), wyznaczając klasy częstości nazwisk polskich, używała w odniesieniu do tych najrzadszych jednostek antropimicznych określenia hapaks-, dis- i trisantroponimy.

zakorzenione już w systemie antroponimicznym lub wręcz stanowiące jego trzon. Dobrze wyraża to zestawienie współczynników korelacji między przebiegiem krzywych dla liczby wszystkich typów (wielkości słownika) oraz imion występujących jeden, dwa lub trzy razy oraz imion o frekwencji 10, 20 i 100. (tab. 5).

Tab. 5. Poziom korelacji między wielkością słownika a liczbą typów imion o określonej frekwencji w kolejnych latach z przedziału 1906–2010

	hapakslegomena	dislegomena	trislegomena	$W_{f=10}$	$W_{f=20}$	$W_{f=100}$
im. żeńskie	0,981	0,978	0,967	0,636	0,243	0,239
im. męskie	0,951	0,977	0,963	0,517	0,609	-0,016

Jak widzimy, w przypadku imion o frekwencji 1, 2 lub 3 korelacja z liczbą typów jest niemal pełna ($\rho > 0,95$), natomiast dla wybranych kolejnych klas częstości możemy mówić o korelacji umiarkowanej, słabej lub bardzo słabej.

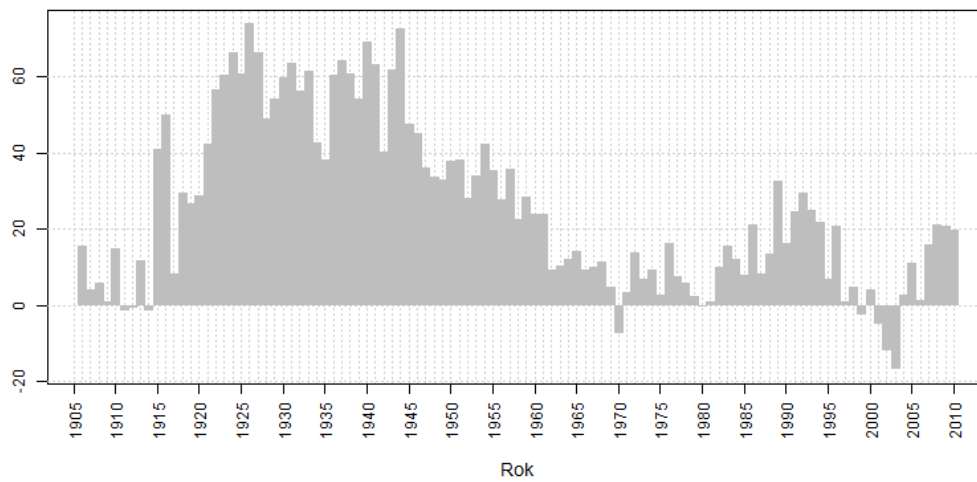
Również same wskaźniki dotyczące udziału najrzadszej leksyki antroponimicznej są ze sobą wzajemnie silnie skorelowane (zob. tab. 6) – zwłaszcza między przebiegiem krzywych dla hapaks- i dislegomenów oraz dis- i trislegomenów występuje niemal pełna zbieżność ($\rho \geq 0,95$), jedynie w przypadku hapaks- i trislegomenów mamy do czynienia z nieco słabszą, choć i tak nadal silną korelacją.

Tab. 6. Poziom korelacji między liczbą hapaks-, dis- i trislegomenów w kolejnych latach z przedziału 1906–2010

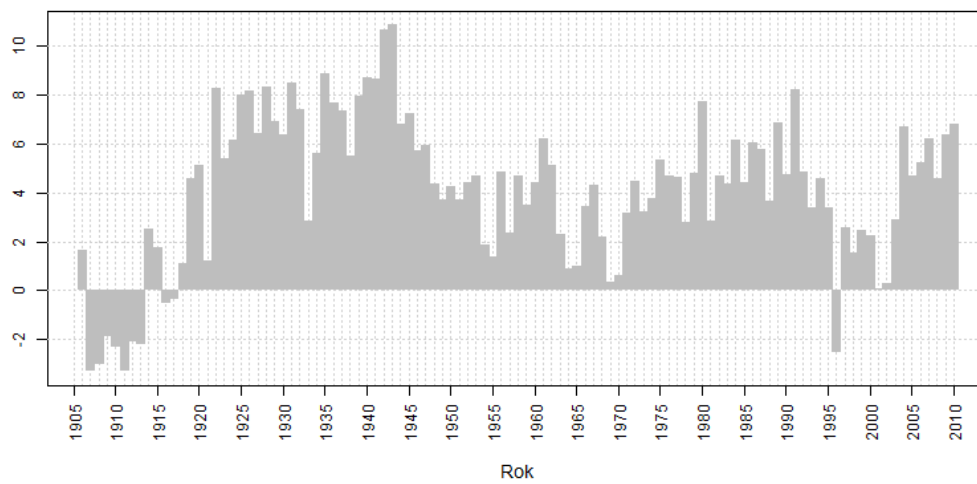
	imiona żeńskie			imiona męskie		
	hapaks.	dis.	tris.	hapaks.	dis.	tris.
hapaks.	1	0,950	0,922	1	0,954	0,870
dis.	0,950	1	0,973	0,954	1	0,951
tris.	0,922	0,973	1	0,870	0,951	1

Biorąc pod uwagę wpływ, jaki na ogólną liczbę typów ma liczba antroponimów o najniższych frekwencjach, nie dziwi zatem fakt, że również w ich przypadku wyższe wartości odnotowano w obrębie imiennictwa żeńskiego, choć zjawisko to nie jest tak regularne, jak w przypadku liczby typów. Zmienność tego zróżnicowania w czasie również wygląda podobnie (rys. 28–30) – także tu mamy do czynienia z intensywnym wzrostem liczby różnic między obiema grupami do lat 40., a następnie z ich stopniowym zmniejszaniem się (zwłaszcza w przypadku hapakslegomenów) aż do końca rozpatrywanego okresu (z chwilowym wzrostem na przełomie lat 80. i 90.).

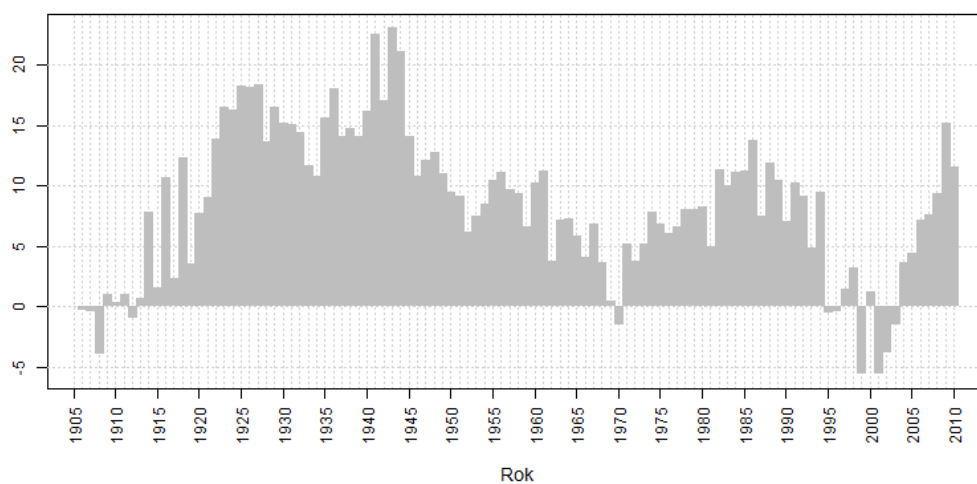
Rys. 28. Różnica w liczbie hapakslegomenów między imiennictwem żeńskim i męskim



Rys. 29. Różnica w liczbie dislegomenów między imiennictwem żeńskim i męskim



Rys. 30. Różnica w liczbie trislegomenów między imiennictwem żeńskim i męskim



B) Liczba hapakslegomenów w imiennictwie na obszarze współczesnych województw (lata 1951–2010)

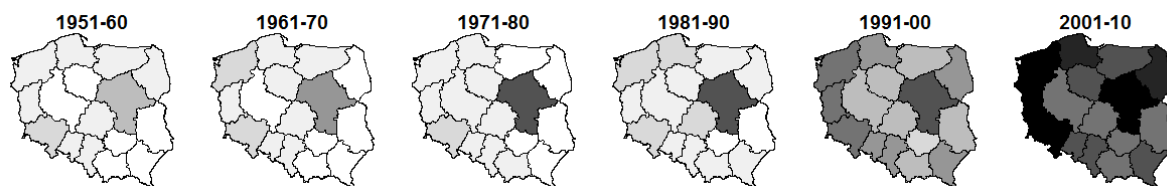
Mając na względzie odnotowany w poprzedniej części wysoki stopień korelacji, jaka zachodzi pomiędzy liczbą hapaks-, dis- i trislegomenów w kolejnych latach, przy omawianiu zróżnicowania regionalnego w zakresie udziału najrzadszej leksyki (tak w skali województw, jak i miast) ograniczę się do przedstawienia wyników obliczeń dotyczących tylko hapakslegomenów.

Tab. 7. Liczba hapakslegomenów na obszarze współczesnych województw

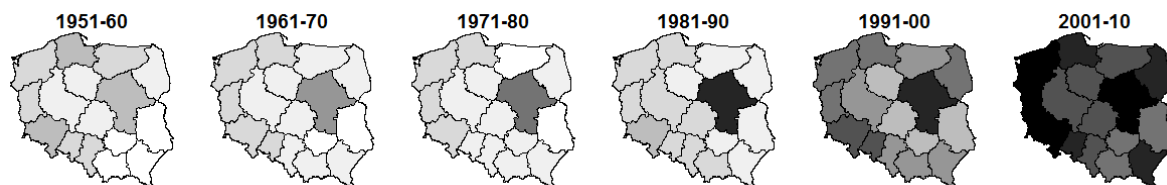
woj.	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	186	164	165	183	302	390	144	158	147	161	281	361
KUJ.	125	128	135	139	195	322	90	106	110	115	204	312
LBL.	101	96	104	116	191	277	79	90	80	98	185	270
LBS.	143	146	153	163	267	379	106	125	114	139	270	396
ŁDZ.	119	124	132	153	212	304	109	122	126	146	207	274
MŁP.	103	114	131	146	215	285	79	98	111	126	208	281
MAZ.	199	237	285	331	356	395	191	250	306	312	323	389
OPL.	163	145	134	144	288	351	136	120	116	115	234	301
PDK.	105	109	109	139	235	325	70	81	96	111	218	307
PDL.	135	114	113	134	261	352	112	106	90	123	246	336
POM.	189	167	168	172	274	347	135	139	135	150	229	325
ŚLS.	145	142	136	148	231	301	120	125	123	129	217	297
ŚWK.	103	97	109	120	203	269	72	84	94	101	178	287
WRM.	149	132	93	128	238	317	122	113	93	112	208	305
WLP.	125	130	136	156	220	298	96	97	109	134	200	285
ZPM.	154	146	160	167	286	392	125	143	138	163	275	371
Śred.	140	137	141	159	248	332	112	122	124	140	230	319
v	22,74	24,74	31,32	31,19	18,21	12,70	28,43	32,98	41,85	35,84	17,09	12,82

Rys. 31. Liczba hapakslegomenów w województwach (dane z tab. 7)

imiona żeńskie



imiona męskie



Układ danych o udziale w imiennictwie poszczególnych województw imion o częstości $f=1$ w tabeli 7 oraz jego kartograficzny obraz (rys. 31) jest poniekąd oczekiwany, zważywszy na fakt bardzo wysokiej korelacji między tym wskaźnikiem a wielkością słownika (III.3.1), jest on bardzo podobny i nie wymaga szerszego komentarza, choć dają się też zauważyć pewne subtelne różnice. Do lat osiemdziesiątych obserwujemy charakterystyczny podział na wschodnią i zachodnią część kraju (z wyraźnie odcinającym się woj. mazowieckim), zaś w ostatnich dwóch dekadach obraz ten ulega już pewnym przekształceniom w postaci coraz wyższej liczby hapaksów w woj. podlaskim oraz podkarpackim (choć tu nadal poniżej średniej).

Nieco inaczej kształtują się różnice w tym zakresie między imiennictwem żeńskim i męskim – choć najmniejsze występowały w województwach centralnych i wschodnich (zwłaszcza w mazowieckim, łódzkim, lubelskim, świętokrzyskim), to niewiele od nich odstaje woj. zachodniopomorskie czy śląskie. Generalnie, tak jak w skali ogólnopolskiej, zwykle więcej hapakslegomenów notuje się w województwach w imiennictwie żeńskim, jednak istnieją od tej tendencji pewne odstępstwa. W drugiej i trzeciej dekadzie w woj. mazowieckim wyższe wartości odnotowano po stronie imion męskich (podobnie jak miało to miejsce w przypadku wielkości słownika), ale na tym nie koniec – od lat 90. taka odwrotna tendencja wystąpiła w woj. lubuskim (1991–2010), kujawsko-pomorskim (1991–2000) oraz świętokrzyskim (2000–2010). Średnie różnice w liczbie hapakslegomenów między imiennictwem żeńskim i męskim na przestrzeni badanego okresu ulegały zmniejszeniu, jednak nie dotyczy to wszystkich województw – tak jak w przypadku liczby typów, wyraźnie odwrotny trend wystąpił w woj. opolskim i łódzkim, a także w dolnośląskim.

Różnicowanie międzyregionalne (wyrażane wskaźnikiem ν) w pierwszych czterech dekadach było znacząco wyższe w obrębie imiennictwa męskiego, jednak pod koniec rozpatrywanego okresu doszło do wyrównania się jego poziomu między imionami żeńskimi i męskimi (w obu przypadkach obserwujemy stopniową niwelację różnic między województwami), co pokrywa się z wynikami dotyczącymi wielkości słownika.

C) Liczba hapakslegomenów w imiennictwie wybranych miast na tle województw (lata 1951–2010)

Z danych zamieszczonych w tabeli 8 wynika, że w ciągu pierwszych pięciu dekad pod względem nadawania najrzadszych imion największe miasta przewyższały pozostały obszar

województw, w których się znajdują. W ostatniej dekadzie widoczne jest zdecydowane zniwelowanie tych różnic (zwłaszcza w przypadku Wrocławia i Poznania) lub nawet odwrócenie tendencji, jak w przypadku imiennictwa męskiego w Krakowie i woj. małopolskim. Na tym tle wyróżnia się zdecydowanie Warszawa, w której imiona o częstotliwości $f=1$ występują kilkakrotnie częściej niż w pozostałej części woj. mazowieckiego i nawet w ostatnich latach, choć ich liczba spadła w stosunku do poprzedzających dekad, nadal występuje ich w stolicy przeszło 100% więcej niż poza nią.

Tab. 8. Liczba hapakslegomenów w ośrodkach wielkomiejskich na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	174	184	186	209	282	332	149	199	163	193	266	323
dśl.	150	133	128	144	243	330	113	118	108	115	229	295
różnica	24	51	58	65	39	2	36	81	55	78	37	28
Kraków	133	148	145	176	211	250	121	133	139	158	203	220
młp.	81	87	99	104	172	236	52	71	75	86	166	246
różnica	52	61	46	72	39	14	69	62	64	72	37	-26
Warszawa	286	363	389	444	453	425	286	373	423	415	404	414
maz.	88	85	95	103	145	212	71	83	82	88	138	208
różnica	198	278	294	341	308	213	215	290	341	327	266	206
Poznań	147	151	146	189	223	254	123	127	143	156	211	250
włp.	100	105	108	114	173	254	76	76	79	100	156	236
różnica	47	46	38	75	50	0	47	51	64	56	55	14

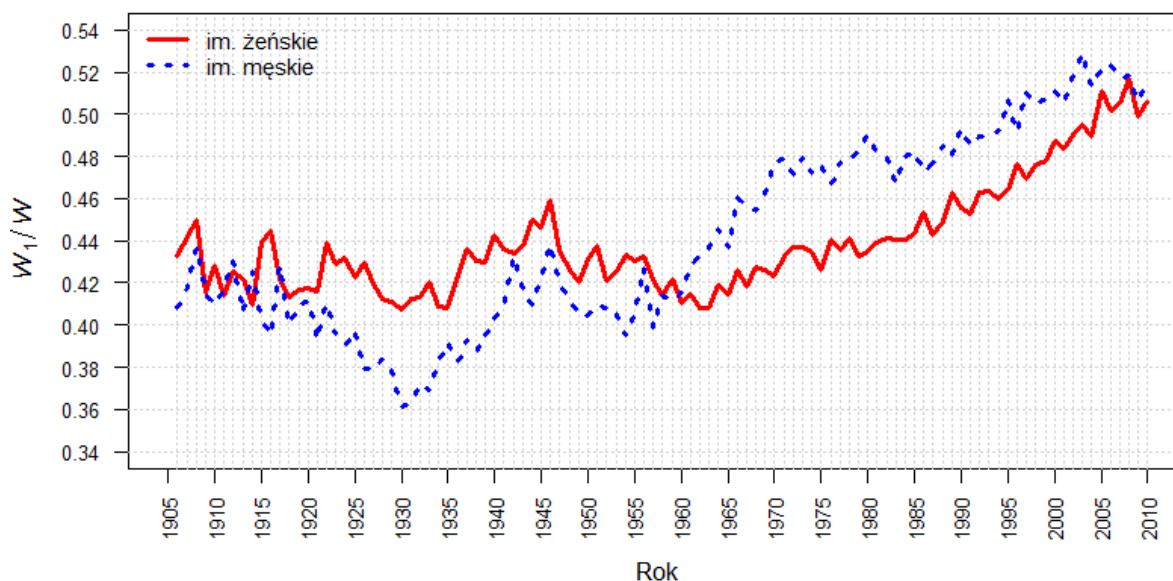
Generalnie zarówno w województwach, jak i ich głównych miastach hapakslegomena częściej występują wśród imion żeńskich, choć istnieją pewne odstępstwa – w Warszawie w latach 1961–80 (tak jak w przypadku wielkości słownika) oraz we Wrocławiu w latach 1961–70 mieliśmy do czynienia z sytuacją odwrotną. Trudno natomiast dopatrzeć się jakichś wyraźnych i regularnych tendencji jeśli chodzi o różnice pod tym względem między województwami i ich stolicami. W przypadku Poznania możemy mówić o na ogół (z wyjątkiem czwartej dekady) mniejszych różnicach w liczbie hapakslegomenów między zasobem imion męskich i żeńskich niż na obszarze woj. wielkopolskiego. W Warszawie i Krakowie taki stan utrzymywał się zwykle do połowy badanego okresu, po czym uległ odwróceniu.

III.3.3. Wskaźnik oryginalności Guirauda

Wskaźnik oryginalności zaproponowany przez P. Guirauda (1954: 34) mierzy stosunek liczby wyrazów o częstotliwości $f = 1$ do wielkości słownika (W_1/W), a więc odsetek liczby haseł słownika, jaki wyczerpują hapakslegomena. Zważywszy na wyniki wcześniejszych analiz składowych tego indykatora (tj. liczby hapakslegomenów i liczby typów), mogłoby się wydawać, że również w jego przypadku będziemy mieli do czynienia z podobnym kształtem wykresu – wszak zachodzi między nimi bardzo wysoka korelacja, przedstawiona uprzednio w tab. 6. Okazuje się tymczasem, że nie jest to takie oczywiste, bowiem przebieg krzywych dla wskaźnika oryginalności Guirauda (rys. 32) już na pierwszy rzut oka odbiega kształtem od wykresów zamieszczonych w poprzednich podrozdziałach.

A) Wskaźnik oryginalności Guirauda w imiennictwie ogólnopolskim (lata 1906–2010)

Rys. 32. Wartości wskaźnika oryginalności Guirauda w imiennictwie ogólnopolskim

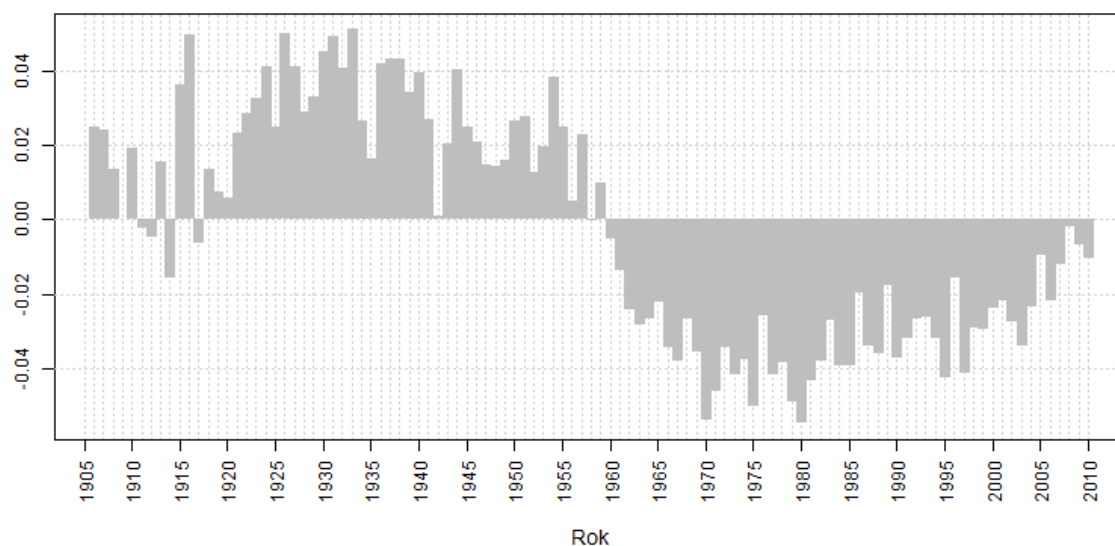


O ile wcześniej mieliśmy do czynienia ze stopniowym wzrostem wartości dla hapakslegomenów i wielkości słownika w okresie do lat 1942–1943, tak teraz widzimy, że w jego pierwszej połowie oryginalność imiennictwa (w ujęciu Guirauda) raczej spadała i dopiero od lat 30. dosyć nieregularnie rosła, a szczyt tego wzrostu przypadł na rok 1946 (w obu grupach). Również spadek wskazań w okresie po II wojnie światowej wygląda odmiennie – jest on tu krótkotrwały – łatwo zaobserwować, że już od przełomu lat 50. i 60.

notowany jest trend rosnący (choć nie możemy tu mówić o ciągłym wzroście) aż do końca rozpatrywanego okresu. Przebieg krzywych dla wskaźnika oryginalności Guirauda tylko częściowo zatem pokrywa się z tymi dla wcześniej badanych miar. Współczynnik korelacji między wskaźnikiem oryginalności a wielkością słownika wyniósł 0,74 (dla imion żeńskich) i 0,5 (dla imion męskich), natomiast przy porównaniu z liczbą hapakslegomenów było to odpowiednio 0,85 i 0,75.

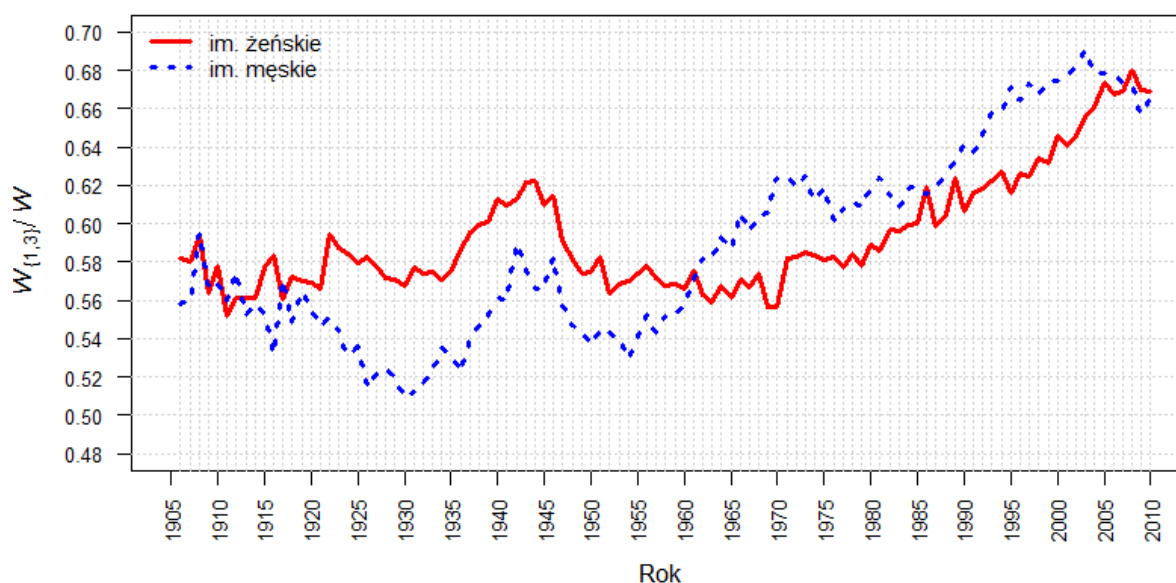
Także różnice między krzywymi dla imiennictwa męskiego i żeńskiego prezentują się zgoła inaczej. Imiennictwo żeńskie odznaczało się większą oryginalnością według wskaźnika Guirauda tylko do końca lat 50., natomiast od roku 1960 notowana jest stała przewaga w tym zakresie imiennictwa męskiego, co obrazuje wykres na rys. 33.

Rys. 33. Różnice w wartościach wskaźnika oryginalności między imiennictwem żeńskim i męskim



Jak już wspomniano, wskaźnik oryginalności Guirauda ujmuje stosunek hapakslegomenów do wielkości słownika. Warto przyrzeć się jednak, czy podobne wyniki otrzymamy w przypadku pozostałych klas najrzadszej leksyki antroponimicznej. W tym celu możemy dokonać jego prostego przekształcenia (analogicznie do modyfikacji wskaźnika koncentracji Mistríka zaproponowanej przez J. Sambor, zob. podrozdział III.3.8), mianowicie zmierzyć udział w całym słowniku antroponimów o częstości od 1 do 3 (a więc $W_{i \leq 3}/W$). Na wykresie na rys. 34 przedstawiono przebieg dla tak zmodyfikowanego wskaźnika oryginalności.

Rys. 34. Wartości zmodyfikowanego wskaźnika oryginalności Guirauda w imiennictwie ogólnopolskim



Jak wynika z przedstawionego wykresu, udział leksyki o częstościach $f = 1, 2, 3$ w całym słowniku prezentuje się podobnie jak w przypadku udziału hapakslegomenów. Widzimy zatem, że nie zawsze bezwzględna liczba tych najrzadszych okazów musi świadczyć o większej różnorodności imiennictwa, jak mogło się wydawać, gdy obserwowaliśmy przebieg krzywych dla hapaks-, dis- i trislegomenów. Należy więc zachować ostrożność w formułowaniu wniosków o przewadze w zakresie bogactwa leksykalnego imiennictwa żeńskiego nad męskim. Nie zawsze też wzrost liczby typów musi pokrywać się ze wzrostem innych wskaźników, co można było zauważyć porównując wskaźnik oryginalności Guirauda z analizowanymi dotychczas wskaźnikami.

B) Wskaźnik oryginalności Guirauda na obszarze współczesnych województw (lata 1951–2010)

Podobnie jak miało to miejsce w przypadku wskaźników dotyczących wielkości słownika oraz najrzadszej leksyki antroponimicznej, również i w przypadku wskaźnika oryginalności Guirauda na przestrzeni całego badanego okresu ewidentnie wyróżnia się województwo mazowieckie, gdzie osiągał on najwyższe wartości i, poza pierwszą dekadą, wynosiły one powyżej 0,5. Za woj. mazowieckim w pierwszych dwóch dekadach plasowały się przede wszystkim województwa z pasa zachodniego, tj. dolnośląskie, pomorskie, zachodniopomorskie, lubuskie i wielkopolskie, a w przypadku imion męskich także łódzkie

(są to obszary, na których wartość wskaźnika była wyższa od średniej). Od 3–4 dekady taki wyraźny podział zaczyna się zacierać i choć nadal w zachodnich województwach mamy do czynienia z wysokim poziomem wskaźnika oryginalności, to często dorównują im obszary wschodnie, jak małopolskie, podlaskie, świętokrzyskie (natomiast, co ciekawe, stosunkowo niskie wartości utrzymują się w woj. śląskim i opolskim, odwrotnie niż w przypadku wielkości słownika). Wiąże się to, jak wynika z obserwacji współczynnika zmienności v , ze stopniowym zacieraniem zróżnicowania międzyregionalnego (od początkowych różnic rzędu ok. 10,5–13,5 do końcowych 3,6–3,8), co dobrze obrazuje fakt, że o ile wcześniej wartość wskaźnika oryginalności powyżej 0,5 notowało tylko województwo mazowieckie, to w dekadzie 2001–10 pułap ten osiągnęło w przypadku imion męskich już 9 województw.

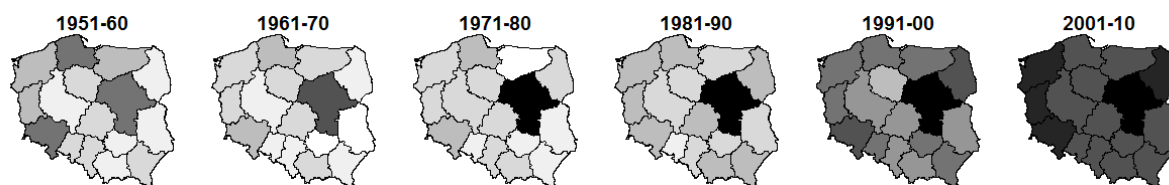
Tab. 9. Wartości wskaźnika oryginalności na obszarze współczesnych województw

woj.	imiona żeńskie						imiona 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.	0,46	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.	0,35	0,37	0,39	0,41	0,44	0,47	0,34	0,38	0,39	0,41	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
OPL.	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
ŚLS.	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
Śred.	0,39	0,38	0,38	0,40	0,45	0,49	0,38	0,39	0,40	0,41	0,48	0,50
v	10,35	10,48	11,84	9,58	6,17	3,63	12,94	13,34	13,03	10,09	4,95	3,87

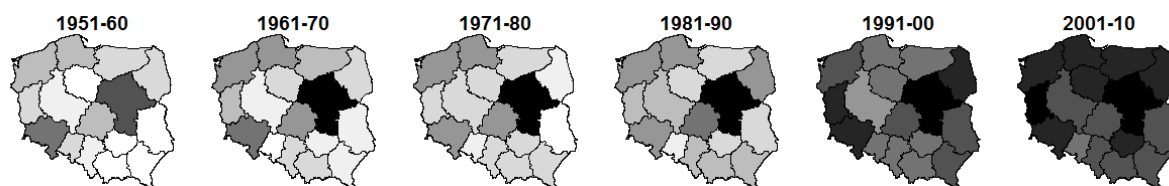
Porównując dane dla województw z tendencjami ogólnopolskimi widocznymi zarówno na wykresie z rys. 34, jak i obrazowanymi przez średnią (tab. 9), możemy zauważyć, że w pierwszej dekadzie, gdy jeszcze utrzymywała się ogólna przewaga w zakresie wskaźnika oryginalności na korzyść imion żeńskich, w województwach łódzkim, podlaskim i mazowieckim doszło już do odwrócenia tej sytuacji. Najwolniej proces ten zachodził z kolei w woj. warmińsko-mazurskim, podkarpackim i opolskim. W tym ostatnim różnice w drugiej dekadzie nie tylko nie uległy zmniejszeniu, ale zwiększyły się o 3%, zresztą poza okresem 1971–80 w woj. opolskim imiennictwo żeńskie stale odznaczało się wyższymi wartościami omawianego wskaźnika, podczas gdy wyniki dla pozostałych obszarów były bliższe tendencjom ogólnokrajowym.

Rys. 35. Wartości wskaźnika oryginalności Guirauda na obszarze współczesnych województw (dane z tab. 9).

imiona żeńskie



imiona męskie



C) Wskaźnik oryginalności Guirauda w imiennictwie wybranych miast na tle województw (lata 1951–2010)

Tab. 10. Wartości wskaźnika oryginalności w ośrodkach wielkomiejskich na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	0,46	0,46	0,46	0,46	0,47	0,50	0,46	0,53	0,46	0,47	0,49	0,51
dśl.	0,43	0,39	0,37	0,39	0,46	0,50	0,41	0,42	0,39	0,39	0,50	0,49
różnica	0,03	0,07	0,09	0,07	0,01	0,00	0,05	0,11	0,07	0,08	-0,01	0,02
Kraków	0,43	0,44	0,43	0,45	0,46	0,47	0,44	0,45	0,46	0,46	0,48	0,46
młp.	0,33	0,33	0,35	0,36	0,43	0,46	0,27	0,33	0,33	0,35	0,45	0,48
różnica	0,10	0,11	0,08	0,09	0,03	0,01	0,17	0,12	0,13	0,11	0,03	-0,02
Warszawa	0,52	0,56	0,55	0,55	0,54	0,53	0,55	0,57	0,56	0,53	0,54	0,54
maz.	0,33	0,32	0,36	0,38	0,42	0,47	0,31	0,36	0,37	0,38	0,45	0,49
różnica	0,19	0,24	0,19	0,17	0,12	0,06	0,24	0,21	0,19	0,15	0,09	0,05
Poznań	0,42	0,40	0,40	0,44	0,44	0,47	0,41	0,42	0,44	0,44	0,45	0,47
włp.	0,34	0,34	0,34	0,35	0,39	0,47	0,31	0,31	0,32	0,37	0,43	0,48
różnica	0,08	0,06	0,06	0,09	0,05	0,00	0,10	0,11	0,12	0,07	0,02	-0,01

Z danych zamieszczonych w tab. 10 wynika, że spośród wszystkich badanych miast Warszawa cechuje się nie tylko najwyższymi wartościami wskaźnika oryginalności, ale także najbardziej odbija się na tle województwa, w którym się znajduje. Najmniejsze wskazania notowano natomiast w Poznaniu. Również tam, oraz we Wrocławiu, odnotowano najmniejsze

różnice między miastem a województwem. Choć w obrębie wszystkich par obszarów na przestrzeni całego badanego okresu doszło do znacznego zmniejszenia lub zaniku różnic pomiędzy imiennictwem miast i prowincji, to w przypadku Warszawy są one większe niż w pozostałych miastach. Łatwo też zauważyć, że wszędzie różnice te były większe w obrębie imiennictwa męskiego, choć w ostatniej dekadzie zaczęło to ulegać zmianie. Nie stwierdzono istotnych różnic między badanymi miastami a województwami w poziomie różnorodności imion męskich i żeńskich.

III.3.4. Wskaźnik różnorodności K Yule'a

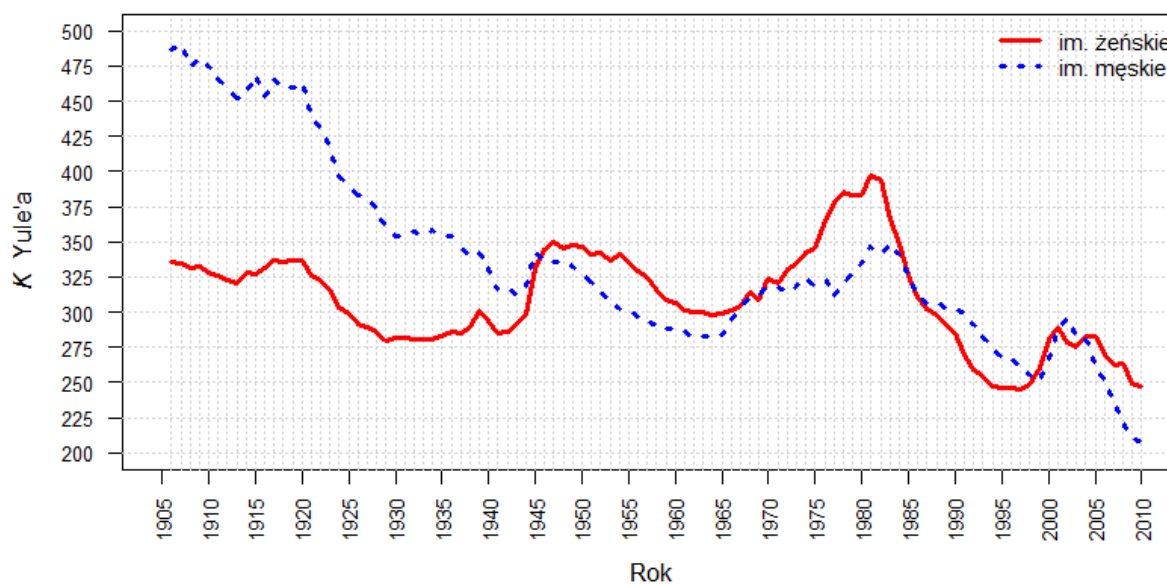
Do oceny różnorodności słownictwa używany jest często wskaźnik K Yule'a (zwany też wskaźnikiem powtarzalności wyrazów, por. Miranda-García, Calle-Martín 2005: 287–288, Huschka, Wagner 2010: 27), który oblicza się na podstawie list frekwencyjnych wyrazów (imion) przy pomocy wzoru (za: Sambor 1969: 158):

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

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). Im wyższa wartość wskaźnika K , tym niższa różnorodność słownictwa w tekście. Jak pisze I. Kamińska-Szmaj, powołując się na J. Woroncza, tekst o większym słowniku (liczbie użytych wyrazów) może posiadać w świetle wskaźnika K mniej różnorodne słownictwo niż tekst o uboższym ilościowo słowniku (Kamińska-Szmaj 1990: 14). Porównanie przebiegu krzywych dla wartości wskaźnika K na linii czasu z krzywymi dla wielkości słownika pozwoli nam na ocenę, czy taki brak zależności ujawnia się także w przypadku materiału imienniczego.

A) Wskaźnik K Yule'a w imiennictwie ogólnopolskim (lata 1906–2010)

Rys. 35. Wartości wskaźnika K Yule'a w imiennictwie ogólnopolskim

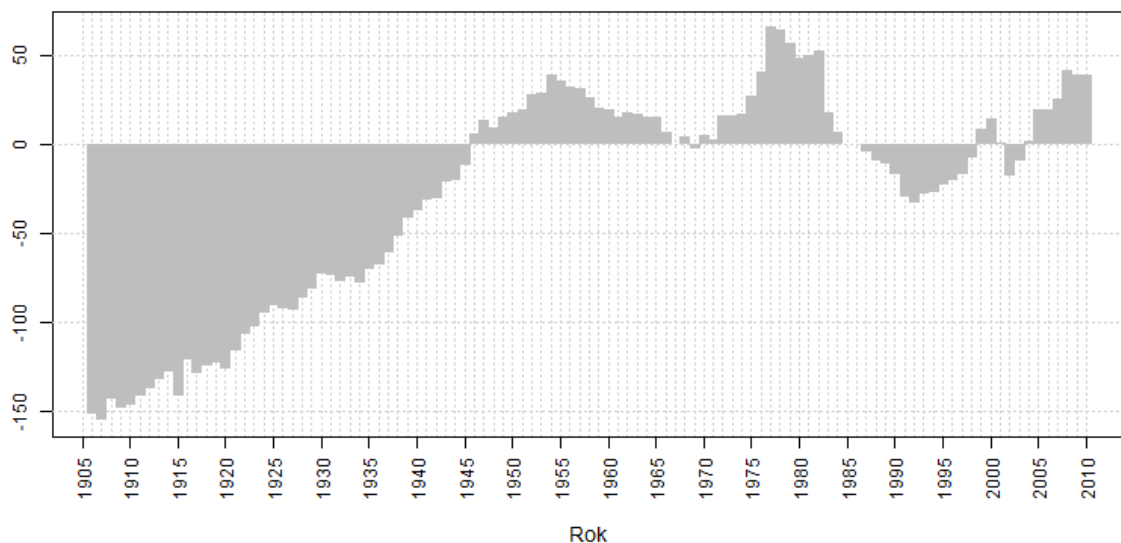


Na podstawie wykresu na rys. 35 możemy stwierdzić, że największe zmiany w zakresie wskaźnika K Yule'a widoczne są w przypadku imiennictwa męskiego. Na początku okresu

wartości K były bardzo wysokie (w 1906 r. było to aż 487), by na przestrzeni całego badanego okresu systematycznie maleć (z kilkoma wyjątkami, zwłaszcza na przełomie lat 70. i 80.), aż do najniższej wartości 207 w roku 2010, a zatem możemy stwierdzić, że w obrębie męskiej antroponimii na przestrzeni przeszło stulecia doszło do znacznego wzrostu różnorodności. Żeńskie imiennictwo również wykazuje się w tym zakresie zmiennością, nie są to jednak różnice tego rzędu, co w przypadku imiennictwa męskiego.

Jeśli chodzi o różnice między imiennictwem żeńskim i męskim, to ich stopniowa niwelacja jest widoczna już na powyższym wykresie (rys. 35), możemy ją jednak dokładnie prześledzić przy pomocy wykresu zamieszczonego na rys. 36. Początkowo różnice we wskaźniku K Yule'a wynosiły między poszczególnymi zbiorami ok. 150. Różnica ta stopniowo malała do 1945 r., a następnie utrzymywała się na stosunkowo niskim poziomie. Co ciekawe jednak, o ile początkowo to imiennictwo żeńskie odznaczało się większą różnorodnością, to już po 1945 r. częściej mamy do czynienia z niższymi wartościami K dla imion męskich.

Rys. 36. Różnice w wartościach wskaźnika K Yule'a między imiennictwem żeńskim i męskim



Wspomniałem wcześniej, że większa różnorodność może, jednak nie musi korelować z większą liczbą typów w słowniku. Współczynnik korelacji między K Yule'a a wielkością słownika wyniósł $\rho = -0,75$ dla imion żeńskich i tylko $\rho = -0,32$ dla imion męskich. Zwłaszcza ten drugi wynik potwierdza zatem, że bogactwo słownikowe także w przypadku zbiorów imion nie musi iść w parze z jego różnorodnością (czy też powtarzalnością wyrazów w jego obrębie).

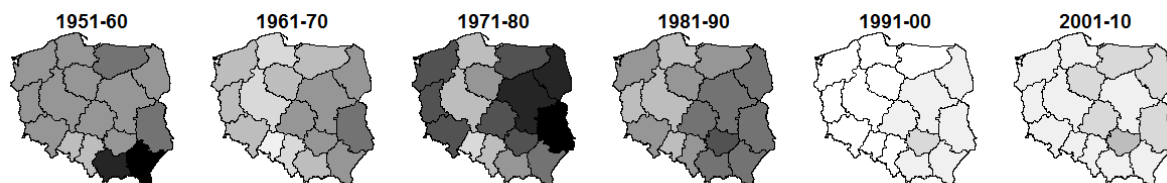
B) Wskaźnik *K* Yule’a na obszarze współczesnych województw (lata 1951–2010)

Tab. 11. Wartości wskaźnika *K* Yule’a na obszarze współczesnych województw

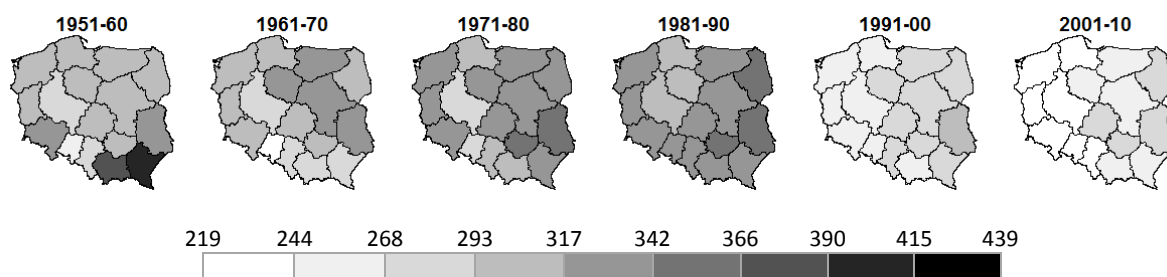
woj.	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	337	314	381	330	234	246	318	302	319	328	261	235
KUJ.	338	301	337	331	241	274	296	321	319	314	269	249
LBL.	353	342	428	364	262	277	325	320	350	346	296	286
LBS.	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
MŁP.	402	307	336	350	264	253	385	291	309	319	266	249
MAZ.	334	334	409	343	257	267	300	321	330	335	271	250
OPL.	302	244	292	303	223	249	262	234	292	322	267	240
PKD.	439	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
ŚLS.	296	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
Śred.	344	304	359	334	245	265	312	299	319	326	271	251
v	9,98	8,66	10,85	6,43	6,10	6,22	12,05	7,50	5,22	4,71	4,97	8,03

Rys. 37. Wartości wskaźnika *K* Yule’a na obszarze współczesnych województw (dane z tab. 11)

imiona żeńskie



imiona męskie



W pierwszej dekadzie najwyższe wartości wskaźnik *K* osiągnął w województwie podkarpackim i małopolskim, najniższe zaś w opolskim i śląskim (w imiennictwie męskim również w warmińsko-mazurskim, natomiast w kolejnych dekadach występuje tendencja, według której wyższe wskazania notowane są na wschodzie Polski – w czołówce utrzymują się niemal zawsze: woj. lubelskie, łódzkie, mazowieckie, świętokrzyskie i podlaskie.

Najniższe wartości notowane były środkowym pasie województw (wyjmując woj. łódzkie), tj. w opolskim, śląskim, warmińsko-mazurskim, wielkopolskim i kujawsko-pomorskim i pomorskim. Choć geograficzna dystrybucja danych nie jest tak wyrazista, jak w przypadku wielkości słownika (najniższe wartości nie dotyczyły województw z zewnętrznego pasa zachodniego), to ewidentnie według wskaźnika K Yule'a na wschodzie Polski imiennictwo (męskie i żeńskie) odznaczało się niższą różnorodnością.

Trudno natomiast dopatrzeć się wyraźnych regionalnych tendencji w zakresie różnic w wartościach K między imionami męskimi i żeńskimi. Z dekady na dekadę geograficzny układ w tym zakresie ulegał ciągłym przeobrażeniom. Uwagę zwraca jednak fakt, że gdy w drugiej dekadzie na obszarze większości województw dokonało się znaczne zmniejszenie różnic lub wręcz odwrócenie tendencji na rzecz wyższych wskazań w obrębie imion męskich (jak w przypadku części województw ze wspomnianego środkowego pasa), to w województwach podkarpackim, lubelskim i małopolskim różnica na korzyść imiennictwa żeńskiego uległa pogłębieniu. I odwrotnie – w ostatniej dekadzie, gdy imiona żeńskie osiągały ponownie wyższe wartości K , to w woj. podkarpackim i lubelskim utrzymywała się przewaga imion męskich (w małopolskim zmalała tylko nieznacznie).

Analiza współczynnika zmienności pokazuje, że na przestrzeni sześciu badanych dekad różnice między województwami w zakresie K Yule'a generalnie ulegały zmniejszeniu, przy czym warto zauważyć, że szybciej proces ten następował w imiennictwie męskim. Co ciekawe, w ostatniej dekadzie obserwujemy, że w przypadku tej grupy imiennictwa doszło do ponownego ich pogłębienia.

C) Wskaźnik K Yule'a w imiennictwie wybranych miast na tle województw (lata 1951–2010)

Z danych w tabeli 12 wynika, że w pierwszej połowie badanego okresu we wszystkich miastach (z wyjątkiem Krakowa w pierwszej dekadzie) wyższe wartości wskaźnika K notowano na obszarze miast, co świadczyć może o niższej różnorodności imiennictwa (odwrotnie niż to miało miejsce w przypadku wielkości słownika). Od czwartej dekady obserwujemy widoczne zmniejszanie tych różnic między miastami i województwami i ostatecznie odwrócenie tendencji w kierunku wyższych wskazań na obszarze poza miastami. Najwcześniej proces ten rozpoczął się w Warszawie, gdzie już w latach 1981–90 wskaźnik K dla imion męskich był o 18 punktów niższy niż w woj. mazowieckim (podczas gdy

w pozostałych miastach był ciągle wyższy o 9–34 pkt.), a dla imion żeńskich tylko nieznacznie, bo jedynie o 2 punkty (na tle 43–62 w innych miastach). Najpóźniej zmiana ta dokonała się w przypadku Krakowa, gdzie dopiero w ostatniej dekadzie zanotowano nieznacznie mniejsze wskazania *K* niż na obszarze woj. małopolskiego (6 pkt. w imiennictwie żeńskim i 8 w męskim).

Tab. 12. Wartości wskaźnika *K* Yule'a w ośrodkach wielkomijskich na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	366	380	445	366	247	235	340	347	338	338	260	226
dśl.	340	305	367	323	233	252	325	298	318	329	265	241
różnica	26	75	78	43	14	-17	15	49	20	9	-5	-15
Kraków	385	363	436	398	276	252	355	324	363	347	278	247
młp.	429	312	314	336	263	258	420	302	299	313	264	255
różnica	-44	51	122	62	13	-6	-65	22	64	34	14	-8
Warszawa	387	422	445	350	253	240	344	380	349	331	255	225
maz.	349	326	406	348	270	305	310	324	338	349	292	292
różnica	38	96	39	2	-17	-65	34	56	11	-18	-37	-67
Poznań	363	338	408	358	236	236	301	351	328	322	257	230
włp.	353	273	292	303	225	257	278	274	287	298	262	250
różnica	10	65	116	55	11	-21	23	77	41	24	-5	-20

Zarówno w wielkich miastach, jak i na poza nimi, różnice w wartościach wskaźnika *K* między imiennictwem żeńskim i męskim zwiększały się (wyższe wskazania w obrębie imion żeńskich) na przestrzeni pierwszych trzech dekad, następnie zaś odnotowano odwrócenie tendencji w tym zakresie. Należy jednak zauważyć, że poziom tych różnic w pierwszych czterech dekadach był na ogół znacznie większy w miastach niż na prowincji, zaś w ostatnich dwu dekadach możemy zaobserwować ich niwelację.

III.3.5. Wskaźnik stereotypowości Mistríka

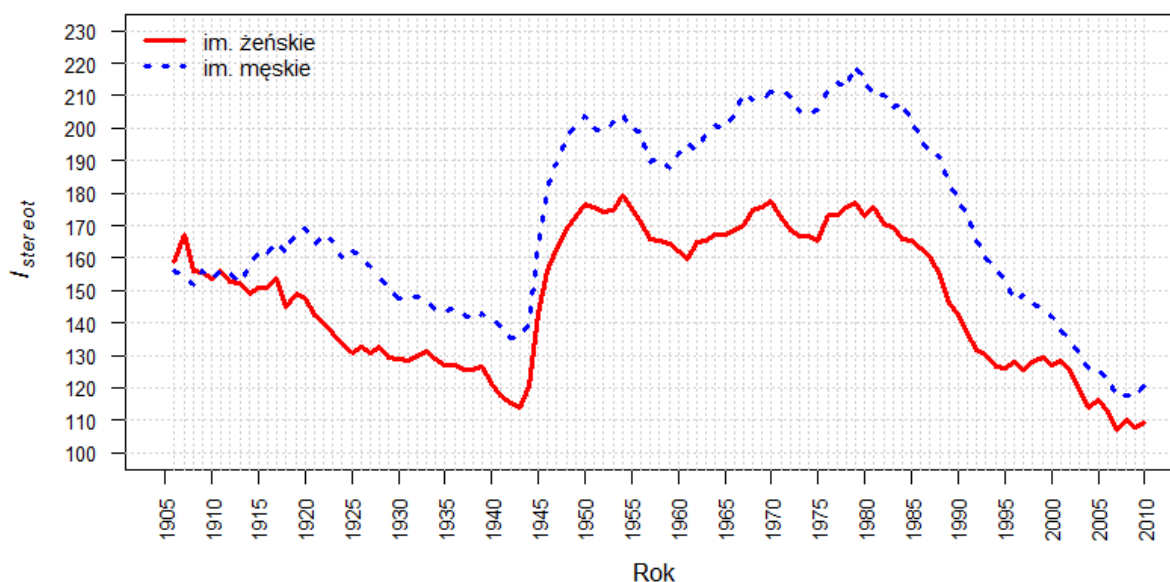
Analizowany w tej części wskaźnik stereotypowości zaproponowany przez J. Mistríka (1967: 45) mierzy przeciętną długość odstępów między dwoma kolejnymi użyciami wyrazów powtarzających się, czyli stosunek liczby typów mających częstości użycia większe niż 1 (Sambor 1972: 223). Jest to stosunek wielkości próby (N) pomniejszonej o liczbę wyrazów o częstości $f=1$ (W_1) do wielkości słownika również pomniejszonego o liczbę hapakslegomenów.⁵⁵ Oblicza się go według następującego wzoru:

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

W badaniach stylistycznych za teksty stereotypowe uznaje się takie, w których odstępy między użyciami wyrazów powtarzających się są duże, natomiast w tekście o słownictwie oryginalnym, zróżnicowanym – krótkie (Mistrík 1967: 45, Kamińska-Szmaj 1990: 17, Gąsiorek 1997: 18, inaczej Sambor 1972: 223).

A) Wskaźnik stereotypowości Mistríka w imiennictwie ogólnopolskim (lata 1906–2010)

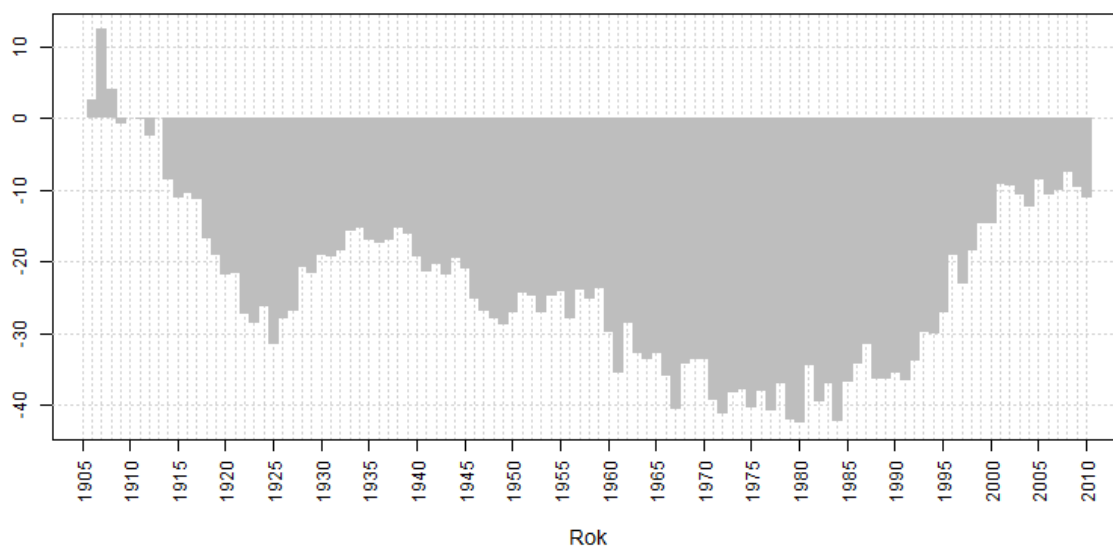
Rys. 38. Wartości wskaźnika stereotypowości Mistríka w imiennictwie ogólnopolskim



⁵⁵ Warto zwrócić uwagę na fakt, iż J. Sambor w swojej pracy podaje prawidłową postać wzoru. W tekście samego Mistríka (1967: 45) pojawia się wzór nieodpowiadający jego opisowi. Podana tam błędna formuła została także powielona w pracy Tuldavy (2005: 375, wzór 8).

Jak wynika z zamieszczonego na rys. 39 wykresu, niemal w całym badanym okresie (z wyjątkiem kilku początkowych lat) imiennictwo męskie odznaczało się wyższym stopniem stereotypowości w ujęciu Mistríka. Warto zwrócić uwagę, że pod koniec ostatniego dziesięciolecia dystans między imiennictwem żeńskim i męskim w zakresie $I_{stereot}$ choć ulegał stopniowej niwelacji, to jednak wydaje się mniej zmienny od różnic w przypadku wielkości słownika.

Rys. 39. Różnice w wartościach wskaźnika stereotypowości między imiennictwem żeńskim i męskim



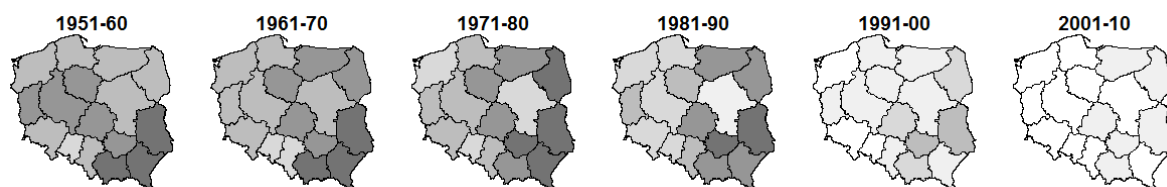
B) Wskaźnik stereotypowości Mistríka na obszarze współczesnych województw (lata 1951–2010)

Tab. 13. Wartości wskaźnika stereotypowości Mistríka na obszarze współczesnych województw

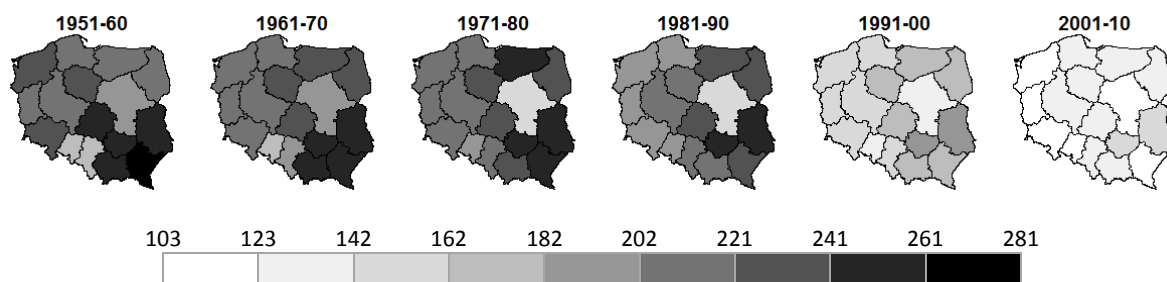
woj.	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	181,6	169,2	162,7	153,9	120,1	102,9	224,2	217,0	208,2	189,8	142,9	112,8
KUJ.	185,6	184,3	185,5	172,2	139,8	118,9	221,4	231,2	232,8	214,5	164,7	129,3
LBL.	205,6	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	148,0	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.	213,8	201,9	194,8	187,2	148,5	125,6	257,9	248,4	230,0	217,3	167,1	131,1
MAZ.	175,7	166,5	156,6	136,7	126,7	112,2	196,2	183,4	161,1	145,1	138,5	119,9
OPL.	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.	220,6	206,7	202,7	188,8	142,3	116,7	280,5	256,9	241,4	233,2	169,1	122,2
PDL.	167,9	193,4	217,3	201,0	143,7	123,6	217,0	233,9	240,2	234,9	168,5	125,2
POM.	173,4	170,2	167,0	157,2	125,9	110,8	212,9	211,2	216,2	196,9	149,1	122,9
ŚLS.	165,1	161,1	173,0	162,6	132,4	118,6	180,1	188,0	207,4	204,4	156,9	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	156,4	119,5	110,0	223,6	214,3	212,1	194,8	143,5	116,3
śred.	185,8	181,2	182,3	174,2	136,6	117,9	223,1	222,2	222,1	210,3	162,0	126,8
v	9,85	9,72	11,09	12,31	11,16	8,00	12,68	11,58	10,79	12,02	11,71	8,17

Rys. 40. Wartości wskaźnika stereotypowości Mistríka na obszarze współczesnych województw (dane z tab. 13)

imiona żeńskie



imiona męskie



Do najwyraźniejszych zjawisk, które można zaobserwować na podstawie danych dla wskaźnika stereotypowości w rozbiciu na województwa w dwóch pierwszych dekadach, należą z pewnością bardzo wysokie wartości dla grupy województw południowo-wschodnich (podkarpackiego, małopolskiego, lubelskiego i świętokrzyskiego), a także bardzo niskie wartości w sąsiadujących ze sobą woj. opolskim i śląskim oraz w woj. mazowieckim. To ostatnie zresztą przez cały badany okres w zakresie imion żeńskich i męskich ewidentnie wyróżnia się na tle pozostałych otaczających wschodnich i centralnych województw. Od 3–4 dekady zaczyna rysować się dosyć wyraźny podział, gdzie niższą stereotypowością odznaczają się województwa zachodniego pasa granicznego (od woj. opolskiego po pomorskie), wyższą zaś województwa wschodnie. W ostatniej dekadzie jednak wyraźnie odstaje w tym względzie woj. podkarpackie, co wpływa na znaczne zmniejszenie zróżnicowania międzyregionalnego wyrażanego wskaźnikiem zmienności $v - z$ ok. 10–12 w pierwszych pięciu dekadach do ok. 8 w dekadzie ostatniej.

Jak mogliśmy zaobserwować wcześniej, poziom stereotypowości w imiennictwie ogólnopolskim niemal zawsze wyższy był w obrębie imiennictwa męskiego i tak samo jest w przypadku wszystkich szesnastu województw. Należy jednak zwrócić uwagę, że różnice między imiennictwem męskim i żeńskim w zakresie wskaźnika stereotypowości były przez pierwszych pięć dekad znacznie niższe w woj. mazowieckim, w pierwszych też w woj. śląskim i opolskim. Przykładowo w latach 1971–80 różnica w woj. mazowieckim wyniosła

tylko 4,5, podczas gdy kolejna różnica – w woj. podlaskim – wyniosła już 22,9, największą zaś odnotowano w woj. zachodniopomorskim – aż 52,4. Natomiast w ostatniej dekadzie zróżnicowanie to uległo niwelacji i nie mamy już do czynienia z takimi rozbieżnościami. Dla przykładu, najniższa odnotowana różnica to 1,6 (woj. podlaskie), najwyższa zaś wyniosła 16,3 (woj. łódzkie). Dobrze obrazuje to korelujący ze spadkiem wskaźnika zmienności spadek średniej różnicy dla wszystkich województw, która w kolejnych dekadach wynosiła odpowiednio: I: 37,3, II: 41,0, III: 39,8, IV: 36,1, V: 25,4, VI: 8,9.

C) Wskaźnik stereotypowości Mistrika w imiennictwie wybranych miast na tle województw (lata 1951–2010)

Tab. 14. Wartości wskaźnika stereotypowości w ośrodkach wielkomiejskich na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	143	140	137	121	94	88	171	167	158	138	107	94
dśl.	153	142	136	132	104	88	186	184	174	165	128	97
różnica	-10	-1	0	-11	-10	0	-15	-17	-16	-27	-21	-3
Kraków	167	161	157	140	119	106	193	187	181	161	136	116
młp.	180	172	164	163	129	108	212	211	194	187	145	110
różnica	-13	-11	-7	-23	-10	-2	-19	-25	-14	-27	-9	5
Warszawa	114	104	94	80	78	80	125	104	89	81	85	84
maz.	166	168	176	176	149	126	192	208	214	210	175	137
różnica	-52	-64	-82	-96	-70	-46	-67	-104	-125	-128	-91	-53
Poznań	147	135	136	122	105	103	166	171	162	149	115	106
włp.	155	145	143	139	112	105	176	179	182	174	145	118
różnica	-8	-11	-7	-17	-7	-2	-10	-8	-20	-25	-29	-12

Spośród badanych miast najniższym wskaźnikiem stereotypowości imiennictwa (męskiego i żeńskiego) odznaczała się Warszawa, zaś najwyższe wartości odnotowano dla Krakowa. Również pomiędzy Warszawą i woj. mazowieckim wystąpiły największe różnice w tym zakresie. Co więcej, o ile w pozostałych miastach poziom zróżnicowania w ostatniej dekadzie był znikomy lub wręcz zerowy (jak w przypadku imiennictwa żeńskiego we Wrocławiu i woj. dolnośląskim), to między Warszawą i pozostałym obszarem woj. mazowieckiego wciąż utrzymuje się spora różnica, sięgająca 46 i 53 punktów. Tłumaczy to zatem odchylenia wskazań dla woj. mazowieckiego, które ujawniły się w badaniu wskaźnika stereotypowości na poziomie województw.

Jeśli jednak mowa o zróżnicowaniu ze względu na płeć, to pod tym względem najmniejsze różnice wystąpiły właśnie w Warszawie – tu najwyższa różnica na korzyść imion męskich wyniosła tylko 11 punktów (w pierwszej dekadzie), podczas gdy w tym samym czasie

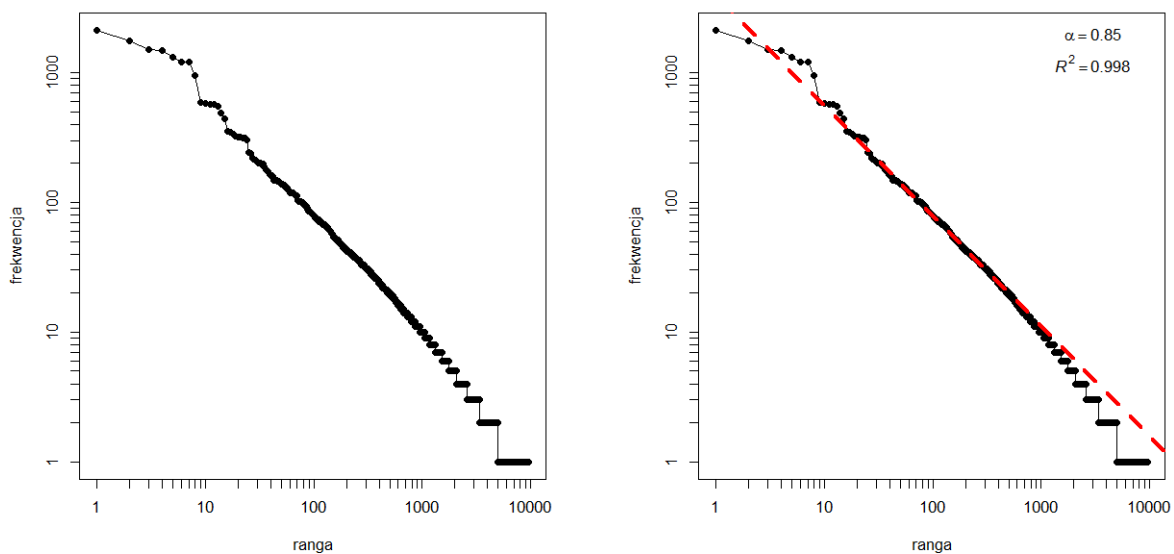
w pozostałych miastach było to od 19 do 28. Również tylko w Warszawie doszło do zrównania pod względem wskaźnika stereotypowości imiennictwa męskiego i żeńskiego, gdy poziom różnic wyniósł 0 (w drugiej dekadzie), a nawet przewagi w imiennictwa żeńskiego o 5 punktów (lata 1961–70). W pozostałych miastach zróżnicowanie między imionami męskimi i żeńskimi wykazywało również tendencję malejącą, ale dopiero w ostatniej dekadzie w Poznaniu osiągnęło poziom o jeden punkt niższy niż w Warszawie. Najwyższy poziom różnic zachował się w Krakowie, gdzie wyniósł 10 punktów. Na ogół różnice między imiennictwem żeńskim i męskim były mniejsze w miastach niż na prowincji (w ostatnich dwu dekadach doszło do nieznacznego odwrócenia tej tendencji w przypadku Krakowa i woj. małopolskiego).

III.3.6. Parametr α

Parametr α z równania Zipfa $f_r \propto r^{-\alpha}$, opisującego zależność między rangą a frekwencją wyrazu, taktowany jest jako miernik różnic w kwantytatywnej strukturze słownictwa tekstu, jednocześnie może być też wykorzystywany jako wskaźnik jego ilościowego bogactwa (Sambor 1972: 70). Określa on kąt nachylenia do osi x krzywej ranga-częstość na wykresie w skali logarytmicznej w przedziale rang, w którym krzywa ta ma przebieg liniowy, tzn. przyjmuje kształt linii prostej. Parametr α ma zatem wartość stałą dla wyrazów o rangach należących do tego przedziału.⁵⁶ W badaniach porównawczych interpretuje się go następująco: im większe wartości α , tym uboższe ilościowo słownictwo.

W celu zobrazowania omawianej liniowej zależności oraz przedstawienia metod obliczania parametru α , posłużę się przykładem wykresu ranga-częstość sporządzonego dla całego tekstu *Pana Tadeusza*⁵⁷. Obserwując krzywą na lewym panelu rys. 41 możemy zauważyć, że w przedziale rang (20,500) ma ona przebieg zbliżony do liniowego.

Rys. 41. Wykres zależności ranga częstość dla tekstu *Pana Tadeusza* A. Mickiewicza. Na prawym panelu czerwoną przerywaną czerwoną linią przerywaną oznaczono prostą dopasowania do modelu.



Aby sprawdzić, czy dane te mają rzeczywiście układ liniowy oraz oszacować dla nich wartość parametru α , w lingwistyce statystycznej stosuje się metodę regresji liniowej (najmniejszych kwadratów), polegającą na dopasowaniu do danych z określonego przedziału rang prostej, która najlepiej charakteryzuje liniową zależność między dwoma zmiennymi,

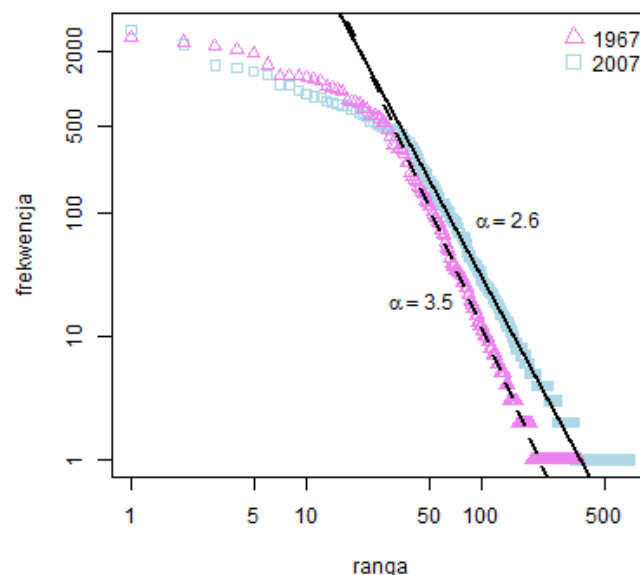
⁵⁶ Zwykle, w celu uproszczenia, zakłada się, że dla danego tekstu $\alpha = \text{constans}$ (Sambor 1972: 70).

⁵⁷ Tekst utworu A. Mickiewicza zaczerpnąłem z portalu wolnelektury.pl. Do jego lematyzacji użyłem usługi LEM (Literary Exploration Machine), opracowanej w ramach konsorcjum CLARIN-PL (<https://ws.clarin-pl.eu/lem>).

czyli rangą i częstością (Baayen 2008: 228, Mehri, Jamaati 2017: 3)⁵⁸. Parametr α stanowi współczynnik regresji (współczynnik kierunkowy/nachylenia, ang. slope) tej prostej. Jakość dopasowania danych z określonego przedziału rang do modelu liniowego ocenia się standardowo przy pomocy współczynnika determinacji R^2 , przyjmującego wartości od 0 do 1 (dopasowanie bardzo dobre zachodzi, gdy $R^2 > 0,9$). W przypadku tekstu *Pana Tadeusza* dla wyrazów rang z przedziału (20, 500) parametr $\alpha = 0,85$, a jakość dopasowania tych danych do prostej regresji wyniosła $R^2 = 0,998$ (zob. prawy panel rys. 41), co oznacza, że wynik ten jest wiarygodny.

W analogiczny sposób w niniejszej części pracy szacowane były wartości parametru α dla zbiorów imion, wykorzystane następnie do porównania tych zbiorów w zakresie bogactwa leksykalnego. Z omówionego powyżej przykładu oraz z badań przytaczanych w rozdziale II.3, wiadomo, że omawiana zależność liniowa dotyczy tylko wyrazów z określonego przedziału rang. Na potrzeby niniejszego badania dla wszystkich podprób imion z każdego roku i regionu sporządzono wykresy ranga-częstość w skali logarytmicznej. Aby obliczyć wartość α należało w każdym badanym zbiorze imion (a więc w każdej podpróbie) niejako „ręcznie”, na podstawie obserwacji wykresów określić przedział rang, w obrębie którego przebieg punktów składających się na krzywe przybierał charakter liniowy. Badanie przeprowadzone na wszystkich podpróbach ujawniło, że w największym stopniu zależność

Rys. 42. Porównanie krzywych ranga częstość dla zbiorów imion męskich z lat 1967 i 2007



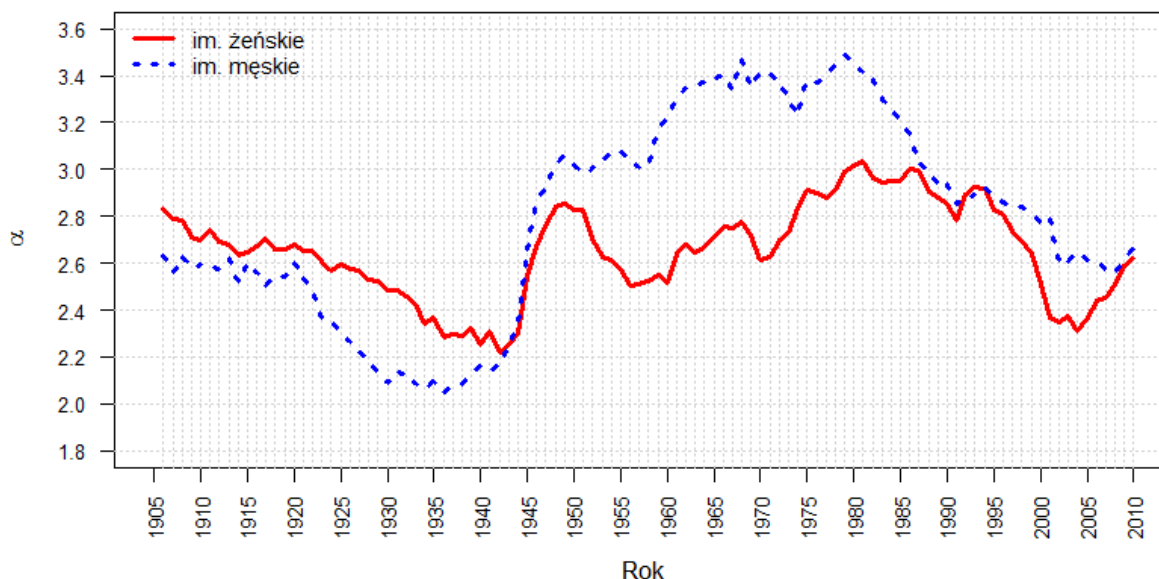
⁵⁸ Zarówno w tym przykładzie, jak i w zasadniczej części badania dotyczącej imiennictwa, dopasowywanie przeprowadzono przy pomocy podstawowej funkcji *lm* (Fitting Linear Models) z pakietu statystycznego R (R Core Team 2013).

liniowa między rangą a frekwencją zachodziła w przypadku imion żeńskich mniej więcej w zakresie rang 50–150, natomiast w przypadku imion męskich 40–140. „Mniej więcej”, gdyż zakresy te mogły się nieznacznie różnić, co zostało uwzględnione. W przypadku wszystkich podprób jakość dopasowania prostej do punktów w danym przedziale rang była wysoka – $R^2 > 0,95$ (w znakomitej większości było to $R^2 \geq 0,99$).

Na rys. 42 przedstawiono przykładowe porównanie wykresów ranga-częstość (w skali logarytmicznej) między zbiorami imion męskich z roku 1967 i 2007 wraz z prostymi dopasowania. W obu przypadkach jakość dopasowania była bardzo wysoka (odpowiednio: $R^2 = 0,997$, $R^2 = 0,998$). Różnica między oboma zbiorami imion jest bardzo wyraźna. Kąt nachylenia linii dopasowania dla roku 1967 (linia przerywana) jest o wiele większy, co odzwierciedla wyższa wartość parametru $\alpha = 3,5$. Możemy na tej podstawie wnioskować, że zasób imion męskich z tego roku był uboższy, mniej różnorodny od imiennictwa z roku 2007. Na wykresie łatwo zresztą zauważyć, że w 1967 r. większy był udział w nominacjach imion najczęstszych (nie licząc pierwszego), zaś w mniejszym stopniu preferowano imiennictwo rzadkie, oryginalne.

A) Parametr α w imiennictwie ogólnopolskim (lata 1906–2010)

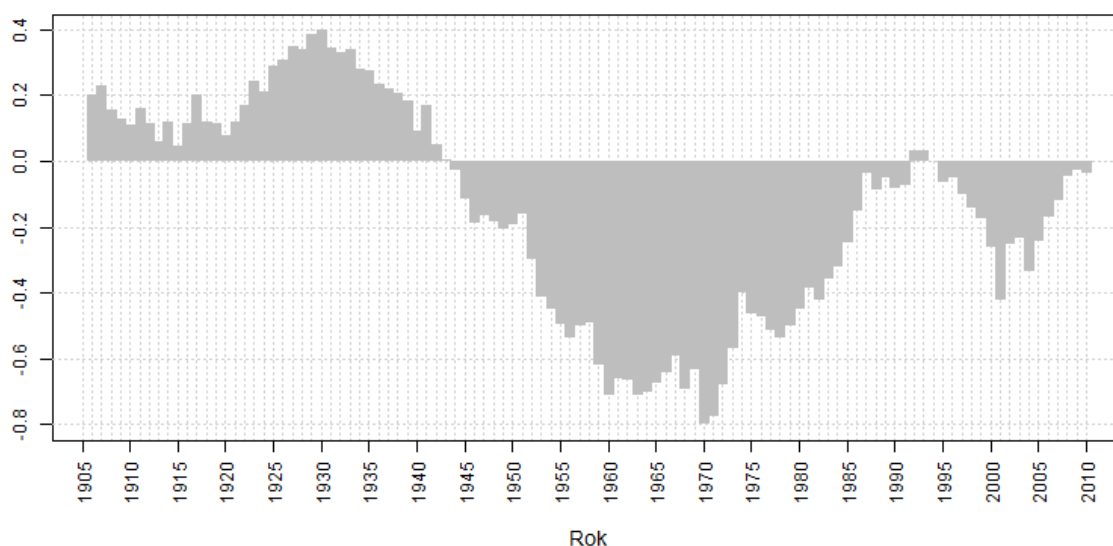
Rys. 43. Wartości parametru α w imiennictwie ogólnopolskim



Wyniki przeprowadzonych w opisany wcześniej sposób obliczeń parametru α dla zbiorów imion męskich i żeńskich w kolejnych latach zostały przedstawione na wykresie na rys. 43. Przebieg krzywych dla parametru α w dużym stopniu przypomina ten dla wskaźnika stereotypowości – współczynnik korelacji między oboma indykatorami był wysoki ($\rho = 0,94$

dla imion żeńskich i $\rho = 0,9$ dla męskich). Również w tym przypadku mamy do czynienia ze względnie regularną tendencją (tu malejącą) od początku badanego okresu do lat 40. XX wieku, a następnie z jej odwróceniem po tym okresie i utrzymywaniem się względnie wysokich wartości (zwłaszcza w przypadku imiennictwa męskiego) do lat 80., od którego to momentu wskazania badanego parametru ponownie zaczynają maleć aż do połowy ostatniej dekady. Co do oceny znaczenia samego parametru α , to imiennictwo polskie generalnie cechuje się jego wysokimi wskazaniami⁵⁹, co świadczy raczej o niskim bogactwie leksykalnym. Jak już wspomniano, w warunkach modelowych, parametr ten przyjmuje wartość 1, natomiast wyższe wartości świadczą o mniejszej różnorodności. W naszym przypadku najmniejsze otrzymane wartości oscylują wokół 2,1, najwyższe zaś osiągają niemalże 3,5.

Rys. 44. Różnice w wartościach parametru α między imiennictwem żeńskim i męskim



Jeśli chodzi o różnice w poziomie parametru α między imionami żeńskimi i męskimi, to ulegała ona znacznym przekształceniom na przestrzeni całego badanego okresu, co obrazuje wykres na rys. 44. Do lat 40. imiennictwo żeńskie cechowało się wyższymi wartościami parametru α , a co za tym idzie uboższą leksyką niż imiennictwo męskie, natomiast w latach późniejszych sytuacja zmieniła się diametralnie i poza nielicznymi przypadkami z połowy lat 90. utrzymywała się do ostatnich lat. Trzeba jednak zaznaczyć, że poziom zróżnicowania zdecydowanie zmalał w stosunku do apogeum, które miało miejsce w latach 60. i 70.

⁵⁹ Przytaczane w rozdziale II.1.3 badania realizacji parametru α w imiennictwie wykazały, że przyjmował on na przestrzeni XX i XXI wartości między 1,3 a 1,9 w USA, zaś w Belgii było to 1,3–2,2 (Li 2012: 46–49, Pomorski i in. 2016: 172–174). Należy mieć jednak na uwadze, że badania te dotyczyły dużo większych zbiorów i przeprowadzenie ich na mniejszych próbach mogłoby dać wyniki nieco odmienne.

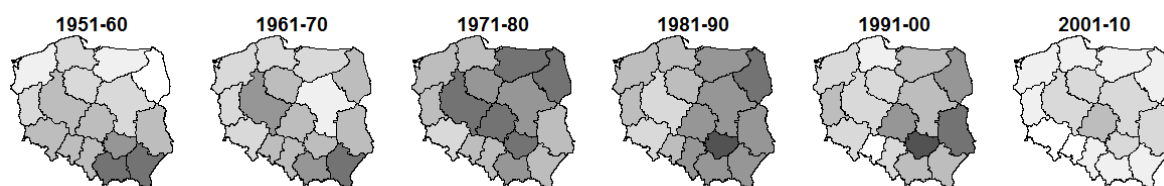
B) Parametr α na obszarze współczesnych województw (lata 1951–2010)

Tab. 15. Wartości parametru α na obszarze współczesnych województw

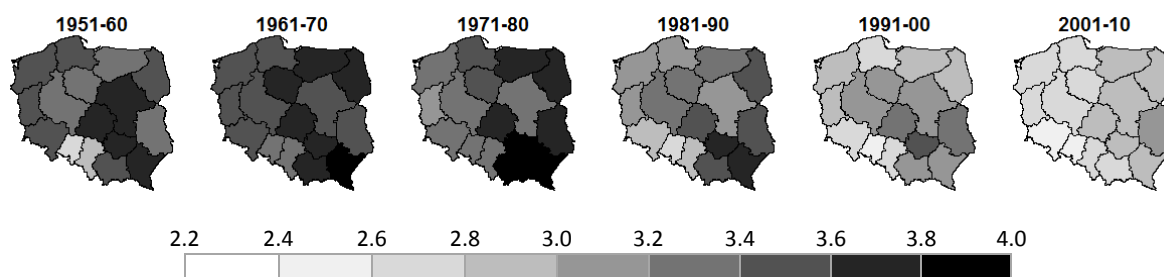
woj.	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	2,85	2,68	2,76	2,76	2,68	2,36	3,55	3,44	3,22	2,81	2,74	2,53
KUJ.	2,63	2,94	2,99	2,90	2,73	2,64	3,34	3,58	3,50	3,37	3,00	2,59
LBL.	2,87	2,92	2,87	3,15	3,36	2,67	3,31	3,56	3,69	3,50	3,29	3,05
LBS.	2,65	2,77	2,77	2,93	2,86	2,47	3,48	3,50	3,14	3,01	2,87	2,66
ŁDZ.	2,80	2,80	3,25	2,99	3,01	2,81	3,65	3,61	3,70	3,56	3,22	2,80
MŁP.	3,30	3,12	3,05	3,04	2,86	2,42	3,41	3,76	3,82	3,51	3,10	2,73
MAZ.	2,69	2,52	3,04	2,98	2,91	2,71	3,58	3,46	3,33	3,08	3,11	2,80
OPL.	2,85	2,86	2,95	2,95	2,26	2,16	2,76	3,28	3,26	2,69	2,42	2,45
PDK.	3,27	3,18	2,85	3,03	2,96	2,42	3,61	3,80	3,98	3,64	3,13	2,77
PDL.	2,35	2,96	3,18	3,33	3,07	2,52	3,53	3,72	3,67	3,48	2,90	2,93
POM.	2,70	2,77	2,91	2,94	2,49	2,38	3,40	3,51	3,47	3,09	2,62	2,62
ŚLS.	2,90	2,92	2,92	3,01	2,61	2,52	2,90	3,26	3,27	2,81	2,59	2,66
ŚWK.	3,02	2,95	3,22	3,42	3,43	2,76	3,59	3,68	3,79	3,61	3,40	2,93
WRM.	2,56	2,69	3,21	3,14	2,96	2,44	3,31	3,58	3,62	3,24	2,96	2,86
WLP.	2,88	2,98	3,25	2,76	2,60	2,62	3,31	3,51	3,33	3,26	3,05	2,57
ZPM.	2,48	2,76	2,78	2,82	2,70	2,38	3,57	3,46	3,21	3,05	2,79	2,67
śred.	2,80	2,86	3,00	3,01	2,84	2,52	3,39	3,54	3,50	3,23	2,95	2,73
v	9,13	5,90	5,91	6,10	10,62	6,92	7,41	4,29	7,37	9,52	9,15	5,96

Rys. 45. Parametr α na obszarze współczesnych województw (dane z tab. 15)

imiona żeńskie



imiona męskie



W odróżnieniu od prezentowanych wcześniej wskaźników geograficzny układ wartości parametru α w pierwszych czterech dekadach jest znacznie zróżnicowany ze względu na płeć osób, którym nadawano imiona. Poza pierwszą dekadą w obrębie imiennictwa męskiego najniższe wartości, świadczące o bogatszej leksyce, notowane były na przestrzeni całego badanego okresu w pasie granicznych województw zachodnich, od śląskiego po pomorskie

oraz w woj. mazowieckim. Do obszarów o wysokich wartościach parametru α zaliczają się z kolei województwa centralne i wschodnie – podkarpackie, małopolskie, świętokrzyskie, łódzkie, lubelskie i podlaskie. W pierwszej dekadzie podział na wschód i zachód nie był tak wyraźny (np. niskie wartości w woj. lubelskim), jednak zdecydowanie najniższe wartości α odnotowano w woj. opolskim i śląskim. W imiennictwie żeńskim taki wyraźny podział, może z wyjątkiem czwartej dekady, jest trudno uchwytny, choć oczywiście i tu z najwyższymi wartościami α mamy do czynienia, tak jak w imiennictwie męskim, w woj. podkarpackim, małopolskim i świętokrzyskim czy łódzkim, najniższe zaś wskazania są zwykle udziałem woj. mazowieckiego i dolnośląskiego, później też opolskiego. Brak jest też zgodności w zakresie wskaźnika zmienności ν – wspólny dla obu grup jest tylko znaczący spadek zróżnicowania międzyregionalnego w drugiej i szóstej dekadzie. W imiennictwie żeńskim ν utrzymywało się na podobnym poziomie od drugiej do czwartej dekady i gwałtownie wzrosło w dekadzie piątej, natomiast w obrębie imion męskich w analogicznym okresie zmienność stopniowo rosła, by w piątej dekadzie zacząć już opadać.

Przyjrzyjmy się teraz temu, jak przedstawiają się dane dotyczące w wartościach parametru α między imionami żeńskimi i męskimi w poszczególnych województwach. Większe wartości notowały, podobnie jak w imiennictwie ogólnopolskim, imiona męskie. W poszczególnych dekadach średnia różnica między imionami męskimi i żeńskimi wynosiła: I: 0,59, II: 0,6, III: 0,5, IV: 0,2, V: 0,1, VI: 0,2 i wykazywała tendencję malejącą, co możemy zauważyć, obserwując stopniowe zmniejszenie intensywności zabarwienia na mapach. Warto jednak zwrócić uwagę, że już w pierwszej dekadzie różnice w α były zerowe lub wręcz odwrotne, jak w woj. opolskim i śląskim, gdzie odnotowano różnice na poziomie odpowiednio 0,09 (o tyle wyższe było α dla imion żeńskich) i 0. Z podobną sytuacją mamy do czynienia w dekadzie czwartej – podczas gdy na obszarze pozostałych województw ciągle wyższe wskazania dotyczą imion męskich, to w opolskim i śląskim α dla imion męskich jest niższe o 0,26 i 0,2 od żeńskich.

C) Parametr α w imiennictwie wybranych miast na tle województw (lata 1951–2010)

Z danych zamieszczonych w tab. 16 wynika, że w pierwszych pięciu dekadach wszystkie badane miasta cechowały się wyższymi wartościami parametru α niż województwa, co świadczy o bogatszej leksyce imiennictwa wielkomiejskiego. Największe różnice, nawet rzędu 1,58 dotyczyły Warszawy i woj. mazowieckiego (im. męskie, 1971–80). W ostatniej

dekadzie różnice między miastami a prowincją są już znikome lub dochodzi do odwrócenia sytuacji (w przypadku imiennictwa żeńskiego bogatszą leksyką odznaczała się prowincja), nie dotyczy to jednak Warszawy, w której nadal parametr α notuje niższe wartości niż w woj. mazowieckim.

W miastach notowano też zwykle częściej mniejsze różnice w parametrze α między imionami męskimi i żeńskimi niż miało to miejsce w województwach, jednak nie możemy tu mówić o pewnym stałym trendzie, a ponadto niejednokrotnie różnice te były znikome.

Tab. 16. Wartości parametru α w ośrodkach wielkomiastowych na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	2,62	2,46	2,66	2,68	2,49	2,54	3,53	3,15	3,04	2,6	2,54	2,55
dśl.	2,84	2,74	2,77	2,74	2,74	2,28	3,5	3,48	3,45	2,87	2,68	2,75
różnica	-0,22	-0,28	-0,11	-0,06	-0,25	0,26	0,03	-0,33	-0,41	-0,27	-0,14	-0,2
Kraków	2,94	2,8	2,87	2,81	2,65	2,57	3,31	3,26	3,48	2,95	2,66	2,74
młp.	3,36	3,22	3,07	3,18	3,02	2,41	3,54	3,77	4,1	3,71	3,1	2,68
różnica	-0,42	-0,42	-0,2	-0,37	-0,37	0,16	-0,23	-0,51	-0,62	-0,76	-0,44	0,06
Warszawa	2,3	2,2	2,55	2,36	2,38	2,55	2,87	2,27	2,38	2,2	2,53	2,77
maz.	2,85	2,71	3,28	3,47	3,44	2,83	3,42	3,57	3,96	3,51	3,12	3,2
różnica	-0,55	-0,51	-0,73	-1,11	-1,06	-0,28	-0,55	-1,3	-1,58	-1,31	-0,59	-0,43
Poznań	2,83	2,83	3,01	2,56	2,35	2,71	3,28	3,09	3	2,9	2,62	2,54
włp.	2,85	3,13	3,28	2,87	2,72	2,59	3,41	3,56	3,63	3,07	2,95	2,75
różnica	-0,02	-0,3	-0,27	-0,31	-0,37	0,12	-0,13	-0,47	-0,63	-0,17	-0,33	-0,21

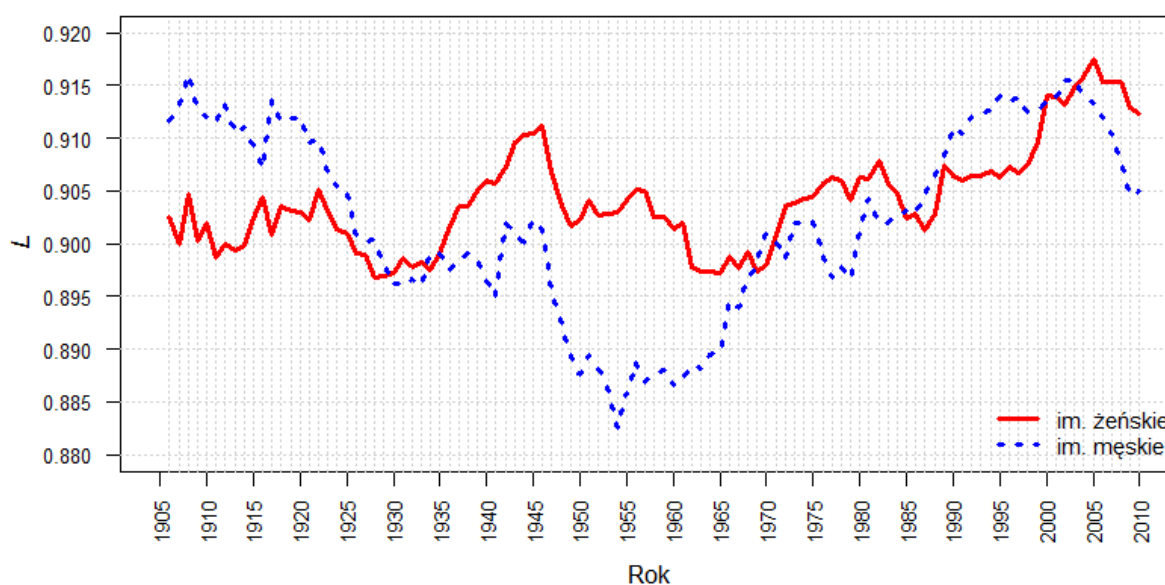
III.3.7. Wskaźnik koncentracji Lorenza

Wskaźnik koncentracji Lorenza, w naukach społecznych i ekonomicznych zwany też wskaźnikiem Giniego, stanowi liczbowy opis nierównomierności podziału ogólnej sumy wartości cechy między poszczególne jednostki zbiorowości (Sobczyk 1998: 56), czyli nierównomierności rozkładu wyrazów w tekście (Hammerl, Sambor 1990: 82). Mierzy on stopień pokrycia tekstu przez częstotliwości r kolejnych wyrazów branych z listy rangowej (Sambor 1969: 113). Ogólnie rzecz ujmując, większy stopień koncentracji imion w danym zbiorze oznacza, że większa jego część 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 (za: Sambor 1969: 158):

$$L = \frac{2 \sum \sum f_i W_i - N}{NW} - 1$$

A) Wskaźnik koncentracji Lorenza w imiennictwie ogólnopolskim (lata 1906–2010)

Rys. 46. Wartości wskaźnika koncentracji Lorenza w imiennictwie ogólnopolskim

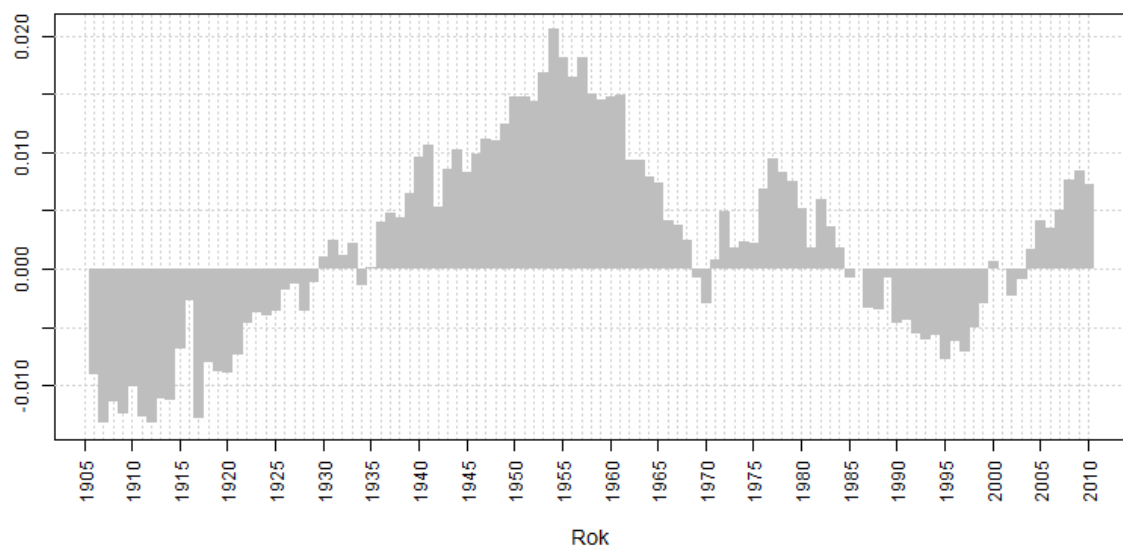


Otrzymane wyniki (przedstawione na wykresie z rys. 46) wskazują na wysoką koncentrację w obrębie polskiego imiennictwa, bo oscylującą wokół 0,9, co pokazuje, że

tylko niewielka liczba najczęstszych imion w dużej mierze wyczerpuje ogólną sumę użyć⁶⁰. Oczywiście wartości L ulegały na przestrzeni czasu wahaniom, co widoczne jest zwłaszcza w przypadku imiennictwa męskiego, którego koncentracja stopniowo malała do przełomu lat 50. i 60. (minimalną wartość – 0,882 – odnotowano w 1954 r.), natomiast po tym momencie mieliśmy do czynienia z tendencją odwrotną. W obrębie imiennictwa żeńskiego zmiany nie są tak gwałtowne, jednak i tu obserwujemy wzrost koncentracji w ostatnich dziesięcioleciach badanego okresu. W obu przypadkach notowany jest również krótkotrwały wzrost wartości wskaźnika koncentracji Lorenza w pierwszej połowie lat 40.

Analizując przebieg krzywych dla wskaźnika koncentracji Lorenza, nietrudno zauważyć, że zwłaszcza w zakresie imiennictwa męskiego odbiega on od uprzednio analizowanych wskaźników bogactwa leksykalnego – najwyższa wartość współczynnika korelacji między L i którymkolwiek wcześniej badanym wskaźnikiem wyniosła zaledwie $\rho = 0,64$ (dotyczyło to porównania z liczbą dislegomenów). Lepsze wyniki odnotowałem w obrębie imiennictwa żeńskiego – tu najwyższa korelacja wystąpiła ze wskaźnikiem oryginalności Guirauda na poziomie $\rho = 0,89$.

Rys. 47. Różnice w wartościach wskaźnika koncentracji Lorenza między imiennictwem męskim i żeńskim



Różnice w wartościach L między imiennictwem żeńskim i męskim, zobrazowane na wykresie na rys. 47 wykazywały zmienność na przestrzeni całego okresu. Do końca lat 20. mieliśmy do czynienia z wyższą koncentracją w obrębie imion męskich, stopniowo jednak

⁶⁰ Podobne wartości (minimum – ok. 0,84 na początku XX w., maksimum – ok. 0,95 pod koniec lat 50.) w przypadku imiennictwa amerykańskiego odnotował W. Li (2012: 45), jednak materiał badawczy obejmował tylko imiona o częstości większej lub równej 5.

zróźnicowanie to malało, a następnie zmienilo swój kierunek i obserwujemy jego znaczne pogłębienie do lat 50. i ponowną niwelację. W kolejnych latach imiennictwo męskie i żeńskie w niemal równych, kilkunastoletnich interwałach naprzemiennie zyskuje przewagę w zakresie koncentracji, przy czym różnice między obiema grupami są niższe niż w pierwszej połowie rozpatrywanego okresu.

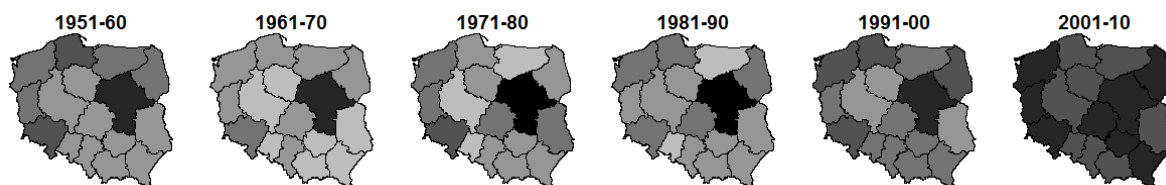
B) Wskaźnik koncentracji Lorenza na obszarze współczesnych województw (lata 1951–2010)

Tab. 17. Wartości wskaźnika koncentracji Lorenza na obszarze współczesnych województw

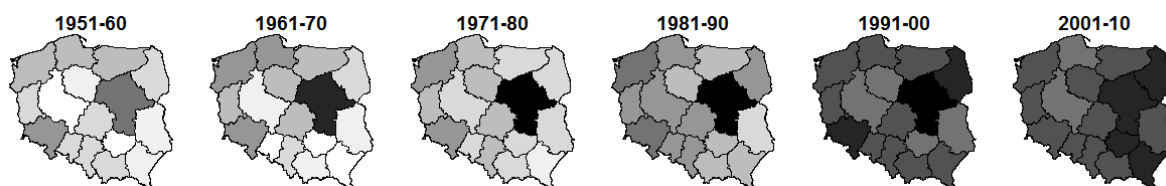
woj.	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	0,908	0,898	0,907	0,903	0,907	0,914	0,890	0,893	0,892	0,901	0,914	0,909
KUJ.	0,894	0,885	0,886	0,892	0,894	0,911	0,864	0,881	0,880	0,882	0,902	0,905
LBL.	0,887	0,882	0,899	0,889	0,894	0,910	0,863	0,868	0,869	0,877	0,902	0,910
LBS.	0,895	0,891	0,898	0,893	0,900	0,913	0,874	0,879	0,880	0,893	0,910	0,910
ŁDZ.	0,893	0,887	0,897	0,899	0,901	0,914	0,870	0,881	0,884	0,893	0,904	0,907
MŁP.	0,890	0,884	0,892	0,894	0,901	0,908	0,871	0,864	0,874	0,884	0,904	0,906
MAZ.	0,913	0,920	0,930	0,927	0,919	0,919	0,902	0,921	0,929	0,929	0,923	0,917
OPL.	0,892	0,881	0,885	0,884	0,900	0,911	0,870	0,860	0,876	0,889	0,907	0,905
PDK.	0,894	0,886	0,888	0,894	0,902	0,913	0,867	0,854	0,865	0,878	0,908	0,913
PDL.	0,897	0,889	0,895	0,897	0,909	0,918	0,871	0,873	0,870	0,890	0,917	0,917
POM.	0,907	0,893	0,894	0,897	0,907	0,911	0,883	0,888	0,883	0,893	0,906	0,902
ŚLS.	0,890	0,886	0,888	0,889	0,898	0,910	0,873	0,876	0,885	0,890	0,909	0,905
ŚWK.	0,887	0,881	0,892	0,887	0,900	0,914	0,851	0,856	0,873	0,880	0,899	0,914
WRM.	0,903	0,893	0,882	0,883	0,900	0,912	0,885	0,881	0,871	0,883	0,906	0,907
WLP.	0,891	0,880	0,883	0,890	0,894	0,905	0,857	0,865	0,872	0,887	0,899	0,901
ZPM.	0,901	0,892	0,901	0,898	0,905	0,916	0,883	0,890	0,888	0,899	0,911	0,909
śred.	0,896	0,889	0,895	0,895	0,902	0,912	0,873	0,877	0,881	0,891	0,908	0,909
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

Rys. 48. Wartości wskaźnika koncentracji Lorenza na obszarze współczesnych województw (dane z tab. 17)

imiona żeńskie



imiona męskie



Najwyższa wartość wskaźnika koncentracji Lorenza w całym badanym okresie występowała zawsze w woj. mazowieckim i dotyczy to zarówno w imiennictwa żeńskiego, jak i męskiego. Kolejne pod tym względem zwykle były woj. dolnośląskie, zachodniopomorskie, pomorskie. Wyniki poniżej średniej notowały z kolei regularnie woj. opolskie i wielkopolskie. To jedyne „stałe” elementy dystrybucji przestrzennej wskaźnika Lorenza – jak łatwo zauważyć na podstawie analizy map, układ ten podlegał przeobrażeniom. Przykładowo wschodnie województwa, jak podkarpackie czy podlaskie, początkowo cechowały się niższą koncentracją w prezentowanym ujęciu, natomiast w ostatnich dekadach wskaźniki *L* należą tam już do najwyższych.

Początkowo zróżnicowanie międzyregionalne, mierzone wskaźnikiem zmienności *v*, ulegało pogłębieniu, jednak już od połowy badanego okresu widoczna jest tendencja niwelacyjna w tym zakresie, przy czym należy zauważyć (w ostatniej dekadzie jest ono najmniejsze), że różnice prawie zawsze były wyższe w imiennictwie męskim.

Jeśli chodzi o różnice w koncentracji między imiennictwem żeńskim i męskim w poszczególnych województwach, to zdecydowanie większe notowane były z reguły na wschodzie Polski. Na przestrzeni całego okresu doszło jednak do ich znacznego zmniejszenia. W pierwszej dekadzie imiona żeńskie miały wskaźnik *L* o 0,023 wyższy, w ostatniej natomiast było to już tylko 0,004.

C) Wskaźnik koncentracji Lorenza imiennictwie wybranych miast na tle województw (lata 1951–2010)

Tab. 18. Wartości wskaźnika koncentracji Lorenza w ośrodkach wielkomiejskich na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	0,900	0,901	0,905	0,905	0,897	0,896	0,884	0,886	0,892	0,898	0,899	0,888
dśl.	0,875	0,865	0,870	0,869	0,886	0,896	0,852	0,839	0,844	0,856	0,887	0,898
różnica	0,025	0,036	0,035	0,036	0,011	0,000	0,032	0,047	0,048	0,042	0,012	-0,010
Kraków	0,899	0,893	0,896	0,899	0,892	0,889	0,874	0,891	0,891	0,895	0,897	0,887
młp.	0,877	0,862	0,864	0,869	0,878	0,895	0,841	0,845	0,848	0,864	0,881	0,889
różnica	0,022	0,031	0,032	0,030	0,014	-0,006	0,033	0,046	0,043	0,031	0,016	-0,002
Warszawa	0,908	0,907	0,915	0,908	0,903	0,901	0,889	0,908	0,898	0,907	0,908	0,897
maz.	0,894	0,882	0,889	0,885	0,892	0,903	0,875	0,871	0,871	0,880	0,901	0,897
różnica	0,014	0,025	0,026	0,023	0,011	-0,002	0,014	0,037	0,027	0,027	0,007	0,000
Poznań	0,926	0,934	0,932	0,925	0,916	0,909	0,919	0,933	0,929	0,926	0,918	0,908
włp.	0,878	0,870	0,883	0,871	0,877	0,898	0,847	0,857	0,859	0,863	0,881	0,893
różnica	0,048	0,064	0,049	0,054	0,039	0,011	0,072	0,076	0,070	0,063	0,037	0,015

Przez pierwszych pięć dekad koncentracja imiennictwa mierzona wskaźnikiem Lorenza w miastach była wyższa niż na prowincji. Największe różnice odnotowano między Poznaniem i woj. wielkopolskim, najmniejsze zaś między Warszawą i woj. mazowieckim (co jest nieco zaskakujące, zważywszy na fakt, iż dotychczas zwykle to właśnie tam występowały największe różnice). Spośród badanych miast, najwyższą koncentracją odznaczał się Poznań, a najniższą Wrocław. Od połowy badanego okresu widoczna jest tendencja polegająca na zmniejszaniu różnic między miastami i obszarem województw. W ostatniej dekadzie różnice te bądź nie występują, bądź większa koncentracja dotyczy imiennictwa na prowincji, jednak w najmniejszym stopniu dotyczy to Wielkopolski.

W przypadku wszystkich porównywanych par obszarów można stwierdzić, że w miastach występowały zwykle mniejsze różnice w koncentracji między imionami żeńskimi i męskimi niż na prowincji. Najmniejsze różnice odnotowano w Poznaniu, największe zaś we Wrocławiu.

III.3.8. Wskaźnik koncentracji Mistríka

W oryginalnym ujęciu J. Mistríka wskaźnik koncentracji mierzy udział w tekście słownictwa o częstościach wyższych niż 1 (Mistrík 1967: 44), a więc całe słownictwo powtarzające się (Sambor 1972: 223). W polskich pracach z zakresu badań stylistycznych (Kamińska-Szmaj 1990: 19; Lewandowska 2008: 75) przyjęto jego zmodyfikowaną przez J. Sambor wersję, która obejmuje leksykę o częstościach wyższych niż 3, gdyż według badaczki do słownictwa „tematycznego”, a więc typowego dla danego stylu tekstu, nie powinno się zaliczać również dis- i trislegomenów (Sambor 1972: 223). Wydaje się zasadne użycie go w takiej postaci także w niniejszej pracy, gdyż jak widzieliśmy w podrozdziale III.3.3, udział w słowniku wyrazów o częstościach 2 i 3 jest silnie skorelowany z udziałem hapakslegomenów. Zmodyfikowany tak wskaźnik koncentracji Mistríka oblicza się według wzoru:

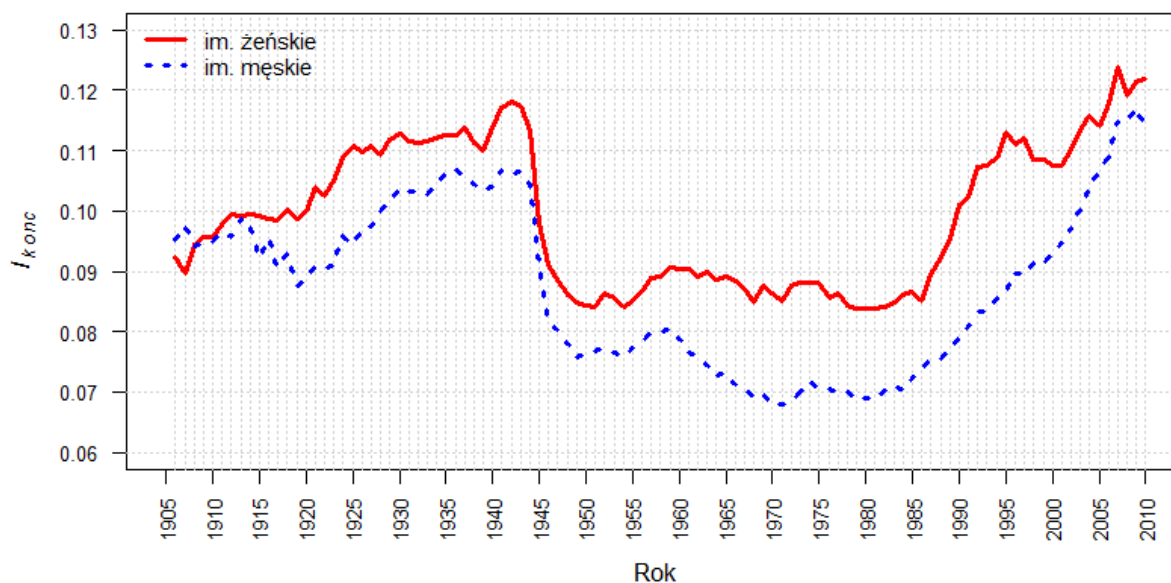
$$I_{konc.} = \frac{20W_{i>3}}{N}$$

Interpretacja wskaźnika I_{konc} na polu stylistyki jest następująca – wyższe wartości wskaźnika świadczą o niższej koncentracji, a więc o większej różnorodności tekstu, natomiast teksty jednorodne tematycznie cechują się jego niskimi wartościami. Na gruncie antroponomii trudno oczywiście mówić o „tematyczności” leksyki, jednak wydaje się, że można go potraktować jako jeden ze wskaźników jej różnorodności, a przede wszystkim skonfrontować go z innymi wskaźnikami koncentracji.

A) Wskaźnik koncentracji Mistríka w imiennictwie ogólnopolskim (lata 1906–2010)

Porównując wyniki dla wskaźnika koncentracji Mistríka zamieszczone na powyższym wykresie (rys. 49) z danymi dla wskaźnika Lorenza (rys. 46), widzimy, że krzywe na obu wykresach mają odmienny przebieg. Jedyna wyraźna zbieżność wystąpiła w latach 40. i 50., gdy obserwujemy gwałtowny spadek poprzedzony krótkotrwałym wzrostem. Ze wzrostem wartości L , który rozpoczął się już w latach 60., wzrost I_{konc} pokrywa się dopiero w połowie lat 80. Korelacja obu wskaźników w czasie jest słaba i wynosi odpowiednio dla imion żeńskich $\rho = 0,37$, a dla imion męskich: $\rho = 0,38$. Widzimy zatem, że w zależności od przyjętej metody mierzenia koncentracji można uzyskać bardzo odmienne wyniki. Różnica między wynikami dla wskaźnika Lorenza i Mistríka wynika z tego, że ten pierwszy bierze

Rys. 49. Wartości wskaźnika koncentracji Mistríka w imiennictwie ogólnopolskim



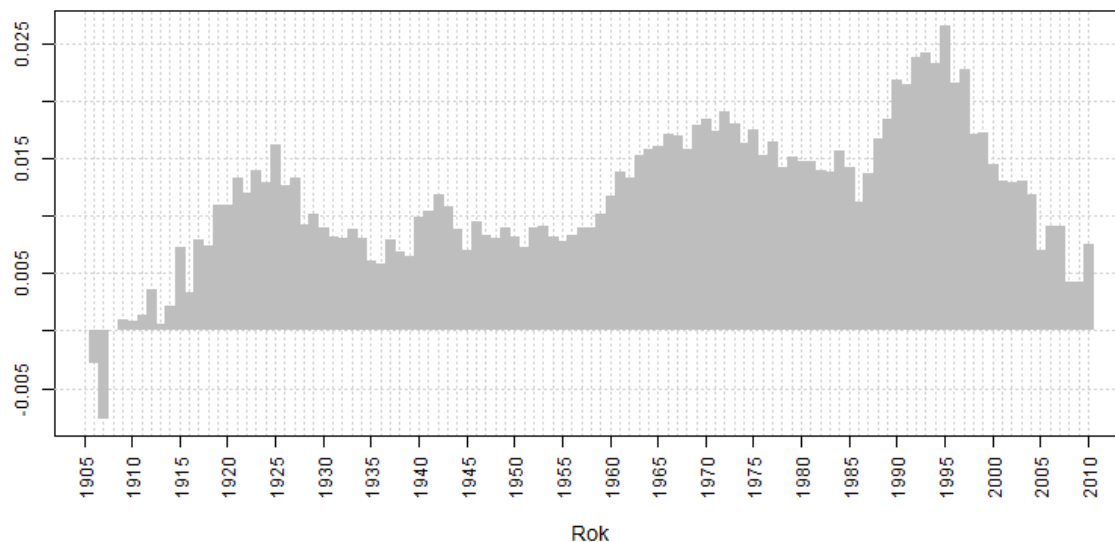
pod uwagę zniuansowanie rozkładu poszczególnych elementów na liście frekwencyjnej, natomiast wskaźnik Mistríka traktuje *en bloc* wszystkie wyrazy o frekwencji powyżej 3 jako częste.

Łatwo zauważyć z kolei podobieństwo do przebiegu krzywych na wykresie dotyczącym wielkości słownika (por. rys. 24) czy też udziału najrzadszej leksyki antroponimicznej (por. rys. 27). W obu przypadkach mamy do czynienia ze stopniowym wzrostem wartości I_{konc} do początku lat 40., a następnie z jego gwałtownym spadkiem i utrzymywaniem się na względnie niskim i stałym poziomie do połowy lat 80., po czym ponownie obserwujemy regularny (zwłaszcza w imiennictwie męskim) wzrost wartości wskaźników aż do końca badanego okresu. Jednak rzeczywistą korelację czasową między I_{konc} a wspomnianymi wskaźnikami tylko w przypadku imion żeńskich możemy określić jako względnie wysoką (wielkość słownika: $\rho = 0,92$, hapakslegomena: $\rho = 0,83$), natomiast dla imion męskich jest dużo niższa (wielkość słownika: $\rho = 0,74$, hapakslegomena: $\rho = 0,5$). Niemal pełną korelację ujemną obserwujemy natomiast przy porównaniu ze wskaźnikiem stereotypowości – tu wyniosła ona odpowiednio: $\rho = -0,99$ i $\rho = -0,96$. Wynik taki właściwie nie jest niczym zaskakującym, gdyż oba wskaźniki zaproponowane przez Mistríka operują podobnymi zmiennymi, tj. słownikiem pomniejszonym o najrzadszą leksykę i wielkością próby (również pomniejszoną o wartość odpowiadającą liczbie najrzadszych leksemów).

Poza kilkoma początkowymi latami w niemal całym rozpatrywanym okresie imiennictwo żeńskie odznaczało się wyższymi wartościami I_{konc} niż imiennictwo męskie. Różnice między grupami ulegały jednak zmianom – na obrazującym je wykresie (rys. 50) możemy zauważyć naprzemienne dłuższe lub krótsze fale wzrostu i niwelacji różnic. Najwyższy poziom różnic

w zakresie I_{konc} między imionami żeńskimi i męskimi odnotowano w połowie lat 90., po czym nastąpił gwałtowny ich spadek.

Rys. 50. Różnice w wartościach wskaźnika koncentracji Mistríka między imiennictwem żeńskim i męskim



B) Wskaźnik koncentracji Mistríka na obszarze współczesnych województw (lata 1951–2010)

Zważywszy na opisane wcześniej zależności pomiędzy wskaźnikami koncentracji i stereotypowości Mistríka, można by oczekiwać, że również w przypadku danych dla I_{konc} w podziale na województwa możemy spodziewać podobnego do $I_{stereot}$ układu przestrzennego danych, oczywiście odwrotnie proporcjonalnego. W dużej mierze tak właśnie jest – jak wynika z danych zamieszczonych w tab. 19, największa koncentracja imiennictwa (a więc najniższe wartości I_{konc}) występowała w województwach wschodnich (poza podlaskim w pierwszej dekadzie i podkarpackim w ostatniej), natomiast najwyższe wartości wskaźnika I_{konc} notowane były na zachodzie Polski, początkowo zwłaszcza na Górnym Śląsku (woj. opolskie i śląskie). Należy jednak zwrócić uwagę na fakt, iż woj. mazowieckie pod względem stopnia koncentracji w mniejszym stopniu odbiegało od pozostałych województw wschodnich – wartość I_{konc} dla tego obszaru zazwyczaj oscylowała wokół średniej. Wyjątek stanowią tu wysokie wyniki w dwóch środkowych dekadach w obrębie imiennictwa męskiego.

Analiza wartości współczynnika zmienności v pokazuje, że zróżnicowanie regionalnie kształtowało się w odmienny sposób w imiennictwie żeńskim i męskim. W obrębie tej pierwszej grupy obserwujemy stopniowy wzrost do 5 dekadę, zaś w imiennictwie męskim

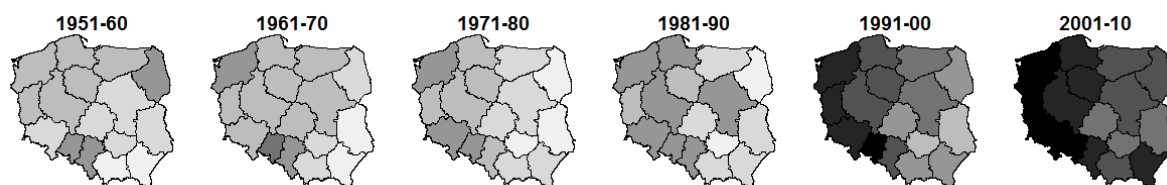
w pierwszych trzech dekad mieliśmy do czynienia ze zmniejszeniem różnic, a następnie z wzrostem w czwartej i piątej dekadzie. Wspólny dla obu grup jest znaczny spadek wartości v między piątą i szóstą dekadą – z ponad 10 do ok. 7,6.

Tab. 19. Wartości wskaźnika koncentracji Mistríka na obszarze współczesnych województw

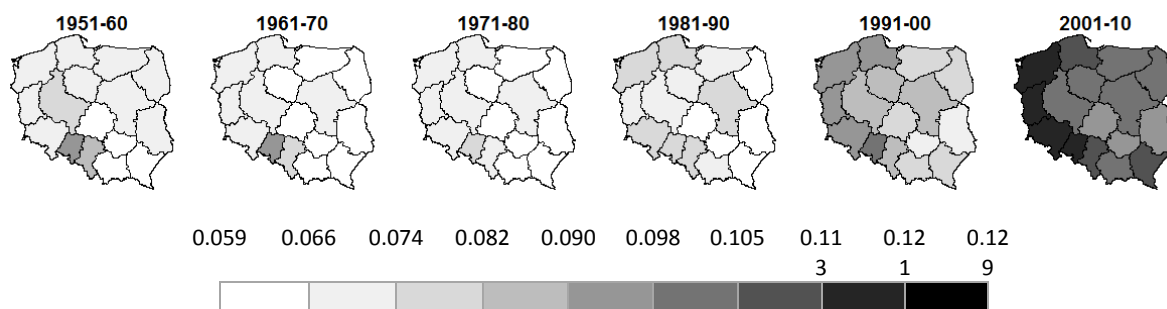
woj.	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	0,081	0,088	0,090	0,092	0,114	0,128	0,068	0,068	0,073	0,078	0,093	0,118
KUJ.	0,083	0,082	0,083	0,087	0,107	0,114	0,071	0,064	0,065	0,069	0,087	0,105
LBL.	0,076	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	0,096
MŁP.	0,072	0,075	0,078	0,078	0,096	0,109	0,064	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
OPL.	0,095	0,099	0,090	0,092	0,123	0,129	0,091	0,090	0,079	0,080	0,102	0,119
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
ŚLS.	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
Śred.	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	8,25	9,17	9,15	9,33	10,65	7,62	11,28	10,76	7,44	8,98	10,18	7,65

Rys. 51. Wartości wskaźnika koncentracji Mistríka na obszarze współczesnych województw (dane z tab. 19)

imiona żeńskie



imiona męskie



Podobnie jak w imiennictwie ogólnopolskim, również w poszczególnych województwach we wszystkich dekadach imiennictwo żeńskie odznaczało się niższą koncentracją niż męskie, choć w ostatniej dekadzie na obszarze w wszystkich województw różnica ta uległa

znacznemu spłyceciu. Zdecydowanie częściej większe różnice między obiema grupami notowane były w województwach skrajnie zachodnich, zwłaszcza zachodniopomorskim i lubuskim, mniejsze zaś występowały na wschodzie Polski (małopolskie, mazowieckie, świętokrzyskie, podlaskie).

C) Wskaźnik koncentracji Mistríka w imiennictwie wybranych miast na tle województw (lata 1951–2010)

Tab. 20. Wartości wskaźnika koncentracji Mistríka w ośrodkach wielkomiejskich na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	0,101	0,105	0,102	0,113	0,140	0,150	0,083	0,084	0,092	0,101	0,117	0,143
dśl.	0,099	0,106	0,109	0,110	0,136	0,150	0,082	0,082	0,089	0,093	0,109	0,137
różnica	0,002	-0,001	-0,007	0,003	0,004	0,000	0,001	0,002	0,003	0,008	0,008	0,006
Kraków	0,085	0,089	0,093	0,098	0,120	0,130	0,077	0,078	0,083	0,089	0,101	0,126
młp.	0,086	0,088	0,094	0,094	0,112	0,130	0,078	0,076	0,078	0,081	0,096	0,122
różnica	-0,001	0,001	-0,001	0,004	0,008	0,000	-0,001	0,002	0,005	0,008	0,005	0,004
Warszawa	0,109	0,110	0,118	0,141	0,148	0,151	0,095	0,110	0,127	0,139	0,140	0,144
maz.	0,093	0,092	0,087	0,087	0,098	0,112	0,081	0,074	0,073	0,072	0,082	0,102
różnica	0,016	0,018	0,031	0,054	0,050	0,039	0,014	0,036	0,054	0,067	0,058	0,042
Poznań	0,099	0,105	0,108	0,122	0,136	0,138	0,090	0,089	0,090	0,092	0,117	0,131
włp.	0,101	0,104	0,107	0,110	0,132	0,132	0,092	0,085	0,086	0,088	0,102	0,119
różnica	-0,002	0,001	0,001	0,012	0,004	0,006	-0,002	0,004	0,004	0,004	0,015	0,012

Spośród badanych miast najwyższą koncentracją wyrażaną przez wskaźnik Mistríka we wszystkich dekadach odznaczała się Warszawa, przy czym należy zaznaczyć, że z czasem różnice między poszczególnymi miastami malały (zwłaszcza Wrocław pod kątem badanego wskaźnika niemal zrównał się z Warszawą). W odróżnieniu od województw we wszystkich miastach występowała stała tendencja rosnąca w zakresie I_{konc} . Warto jednak zauważyć, że tylko Warszawa wyraźnie odróżnia się na tle województwa, w którym leży. W przypadku pozostałych miejscowości różnice te są marginalne.

We wszystkich miastach z reguły również występowały znacznie mniejsze różnice w koncentracji pomiędzy imionami żeńskimi i męskimi. Na tym tle ponownie wyróżnia się Warszawa, w której poza pierwszą dekadą różnice te były najmniejsze. W drugiej dekadzie w ogóle nie odnotowano tam różnic, a w trzeciej doszło nawet do sytuacji nieznacznie większej koncentracji w obrębie imion żeńskich. W ostatniej dekadzie doszło już do znacznego zmniejszenia i wyrównania poziomu różnic we wszystkich miastach.

III.3.9. Pokrycie sumy użyć przez r najczęstszych imion

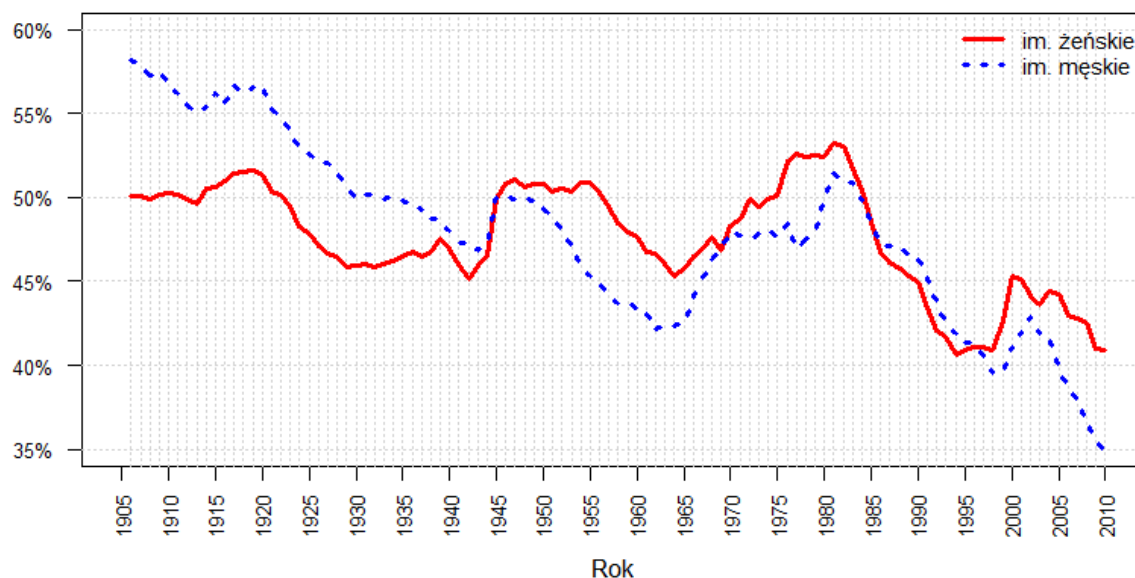
Według S. Liebersona jednym z najbardziej wyrazistych i trwałych zjawisk, jakie dokonały się w zachodnim imiennictwie w XX w. był wielki spadek jego koncentracji (Lieberson 2000: 42–68; Lieberson, Lynn 2003: 238–243). Miara, której użył badacz do jej określenia była bardzo prosta – zbadał on udział arbitralnie określonej liczby najczęstszych (zwykle od 10 do 25) imion w korpusach imienniczych poszczególnych krajów Europy i Ameryki Północnej. Okazało się, że od przełomu XIX i XX w. następuje systematyczny spadek udziału najczęstszych imion (zarówno męskich i żeńskich). Należy przy tym zaznaczyć, że Lieberson obliczał tak ujęty wskaźnik koncentracji prawdopodobnie na bardzo nierównych próbach, a w związku z tym mógł on być uzależniony od ich długości (a właściwie od wielkości słownika, która rośnie wraz z nią). Niemniej jednak wyniki otrzymane przez Liebersona wyraźnie pokazują, że najczęstsze imiona w coraz mniejszym stopniu dominują imiennictwo poszczególnych krajów, a zatem jego koncentracja ulega zmniejszaniu. Miara wykorzystana przez Liebersona jest wygodna i nie wymaga złożonych obliczeń, obarczona jest jednak, jak już wspomniano pewnym ryzykiem. Trudno jest bowiem jednoznacznie określić jaką liczbę imion w danym zbiorze uznać za liczbę najczęstszych imion. Warto, mając na uwadze powyższe wątpliwości, sprawdzić, czy odnotowane przez Liebersona zjawisko wystąpiło również w polskiej antroponimii. Na wykresach z rys. 52 i 53 przedstawiono procentowy udział 10 i 20 najczęstszych imion w ogóle imiennictwa, tj. stosunek sumy nosicieli imion o 10 i 20 najniższych rangach do ogółu użytkowników.

A) Pokrycie sumy użyć przez 10 i 20 najczęstszych imion w imiennictwie ogólnopolskim (lata 1906–2010)

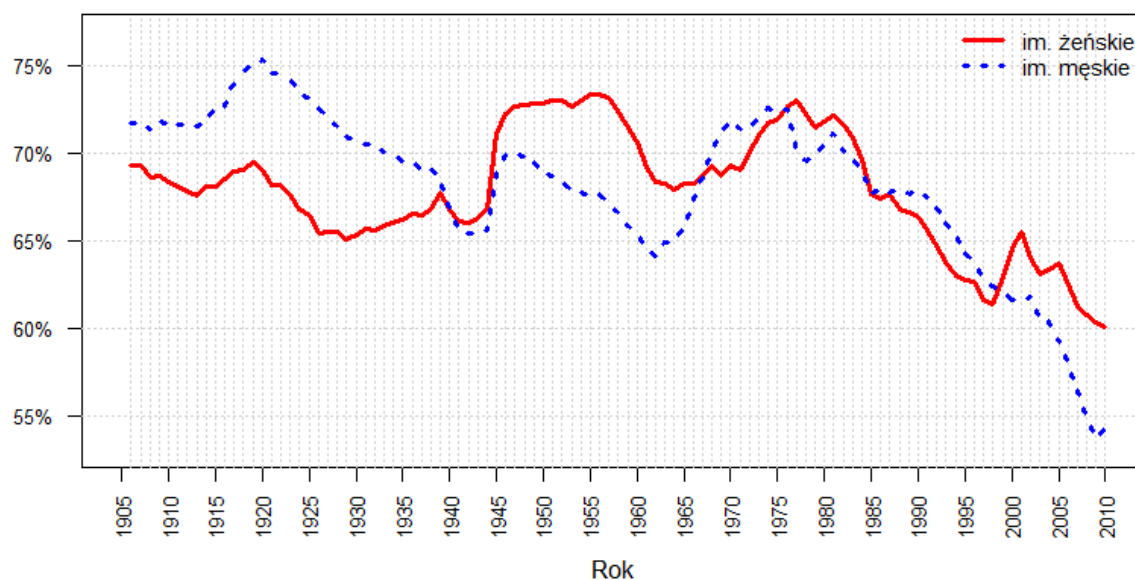
Porównując wyniki dla badanych wskaźników (rys. 52 i 53) z początku i końca rozpatrywanego okresu widzimy, że zwłaszcza w obrębie imiennictwa męskiego nastąpiła duża zmiana w koncentracji w ujęciu Liebersona. W 1906 r. pierwszych dziesięć imion męskich z listy frekwencyjnej nadano ponad 58% osób, a dwadzieścia najczęstszych antroponimów wyczerpało niemal 72% sumy użyć (w 1920 r. było to nawet 75,4%). Pod koniec rozpatrywanego okresu było to już tylko odpowiednio 34,9% i 53,6%, a więc mowa tu o różnicy rzędu ponad 20 punktów procentowych. W przypadku imion żeńskich różnice te mają dużo mniejszą skalę – początkowo 10 najczęstszych imion nadawano ok. 50% wszystkich osób, natomiast w ostatnich latach było to ok. 41%. Udział 20 najczęstszych

imion zmniejszył się z początkowych 69% do 60% w roku 2010. Spadek ten nie dokonywał się jednak płynnie – na wykresach obserwujemy, że był on przerywany dużymi wzrostami koncentracji – po 1945 r., na przełomie lat 60. i 70. czy na początku lat dwutysięcznych. Widoczne jest to przede wszystkim w obrębie imiennictwa żeńskiego.

Rys. 52. Pokrycie sumy użyć przez 10 najczęstszych imion w imiennictwie ogólnopolskim



Rys. 53. Pokrycie sumy użyć przez 20 najczęstszych imion w imiennictwie ogólnopolskim

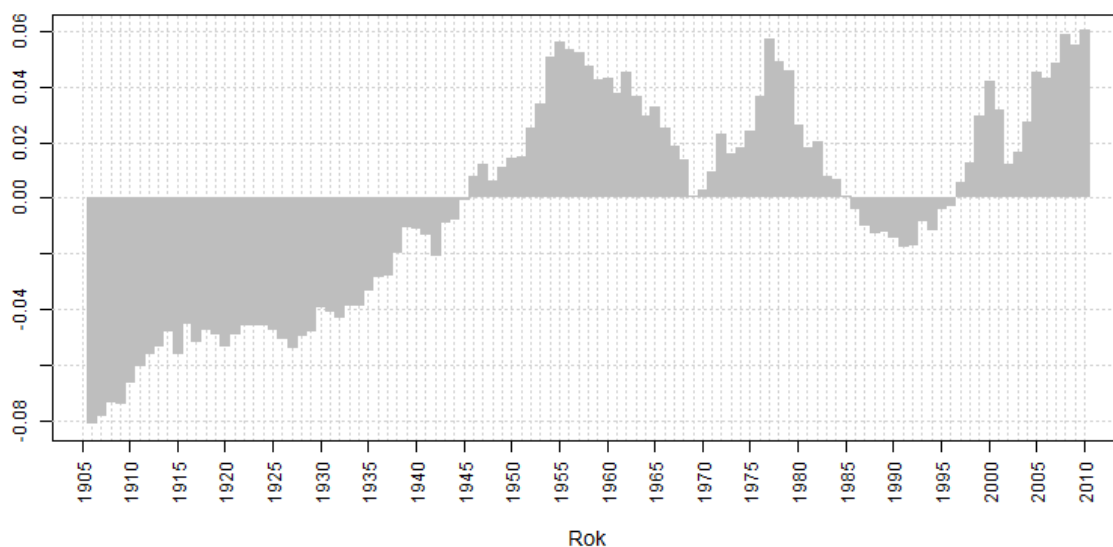


Przyglądając się przebiegowi krzywych na obu wykresach, łatwo zauważyć, że w znacznym stopniu odbiega on od wyników uzyskanych przy analizie wskaźników

koncentracji zarówno Lorenza, jak i Mistríka, co kolejny raz potwierdza, że wskaźniki te nie mogą być stosowane zamiennie.⁶¹

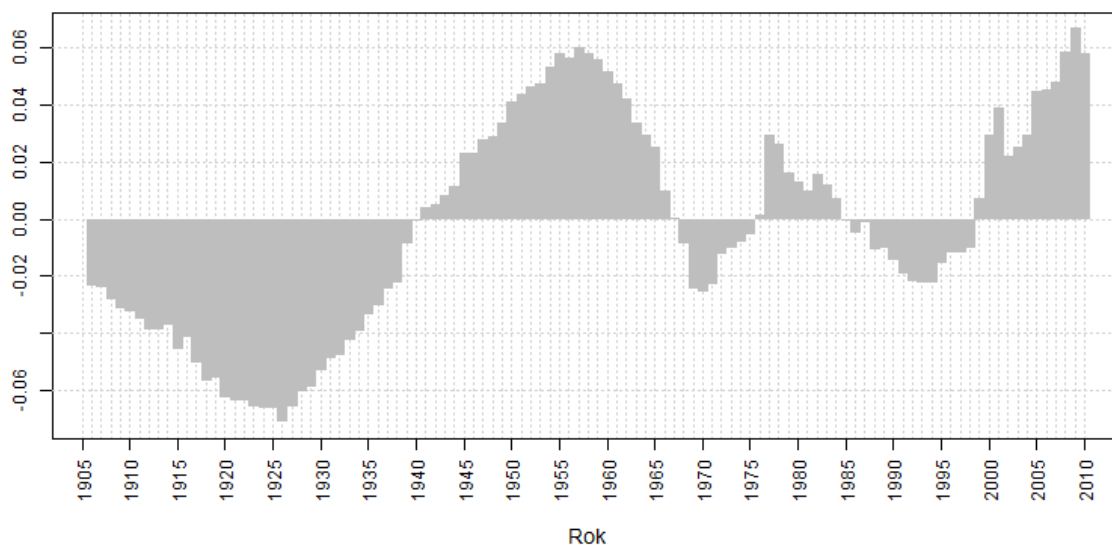
W dużej mierze dane o pokryciu zbioru przez 10 i 20 najczęstszych imion w Polsce przypominają natomiast wyniki, które otrzymałem dla wskaźnika różnorodności *K* Yule'a. Zwłaszcza dane dla 10 najczęstszych imion (zarówno męskich, jak i żeńskich) wykazują bardzo wysoką korelację z tym wskaźnikiem ($\rho = 0,96$). Co ciekawe, jest ona nawet silniejsza niż w przypadku porównania przebiegu krzywych dla 10 i 20 najczęstszych antroponimów (im. żeńskie – 0,88 i im. męskie – 0,89). Świadczy to o tym, że wskaźnik *K* jest w znacznej mierze zależny od frekwencji najczęstszych imion. Ponadto pokazuje, że w zależności od liczby dobranych najczęstszych imion możemy otrzymać nieco odmienne wyniki. Uwidacznia się to także przy porównaniu wykresów dotyczących różnic w wartościach danego wskaźnika między imiennictwem żeńskim i męskim, zobrażowanych na rys. 54 i 55. Przez całe lata 70. w przypadku 10 najczęstszych imion mieliśmy do czynienia z wyższą koncentracją w obrębie imion żeńskich, natomiast gdy spojrzymy na wykres na rys. 55, otrzymamy obraz nieco odmienny – w pierwszej połowie lat 70. to imiona męskie odznaczały się wyższym pokryciem sumy użyć przez 20 najczęstszych imion.

Rys. 54. Różnica w procentowym pokryciu sumy użyć przez 10 najczęstszych imion między imiennictwem żeńskim i męskim



⁶¹ Twierdzenie Liebersona i Lynn o malejącej koncentracji na przestrzeni całego XX w. w imiennictwie USA i innych krajów opiera się wyłącznie na zmierzeniu procentowego udziału najczęstszych imion w ogólnej sumie użyć. Tymczasem badanie koncentracji przeprowadzone przez W. Li (2012: 45–46) na materiale imion amerykańskich przy użyciu wskaźnika Giniego/Lorenza również ukazało inny obraz, mianowicie wzrost koncentracji do końca lat 50. (intensywny zwłaszcza na początku XX w.) i stopniowy spadek w późniejszym okresie.

Rys. 55. Różnica w procentowym pokryciu sumy użyć przez 20 najczęstszych imion między imiennictwem żeńskim i męskim



B) Pokrycie sumy użyć przez 10 i 20 najczęstszych imion w imiennictwie na obszarze współczesnych województw (lata 1951–2010)

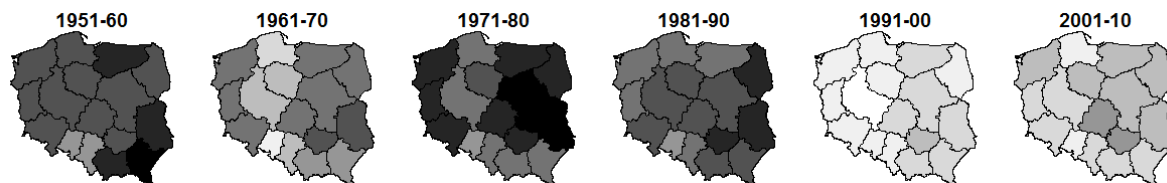
Mając na uwadze wysoką korelację między wskaźnikiem K Yule'a i pokryciem sumy użyć przez 10 najczęstszych imion, nie jest dla nas zaskoczeniem, że również dystrybucja przestrzenna wartości badanych wskaźników jest niemal identyczna (tab. 21, rys. 56), nie ma zatem potrzeby powtarzać w tym miejscu szczegółowych obserwacji. Generalnie przez cały badany okres zdecydowanie większa koncentracja utrzymuje się we wschodnich województwach Polski, zaś niższe wartości notowane były w pasie środkowym, zwłaszcza w woj. śląskim, opolskim, wielkopolskim i pomorskim. Podobieństwa widoczne są także w przypadku współczynnika zmienności – zasadniczo jeśli porównamy wartości v z początku i końca badanego okresu, to zróżnicowanie między regionami w ciągu sześciu dekad uległo zmniejszeniu (jest to widoczne zwłaszcza w imiennictwie męskim), jednak w ostatnich dwóch dekadach nastąpiło jego nieznaczne pogłębienie. W zbliżony sposób do kształtuje się konfiguracja geograficzna w zakresie pokrycia sumy użyć przez 20 najczęstszych imion (tab. 22, rys. 57).

Tab. 21. Pokrycie sumy użyć przez 10 najczęstszych imion na obszarze współczesnych województw

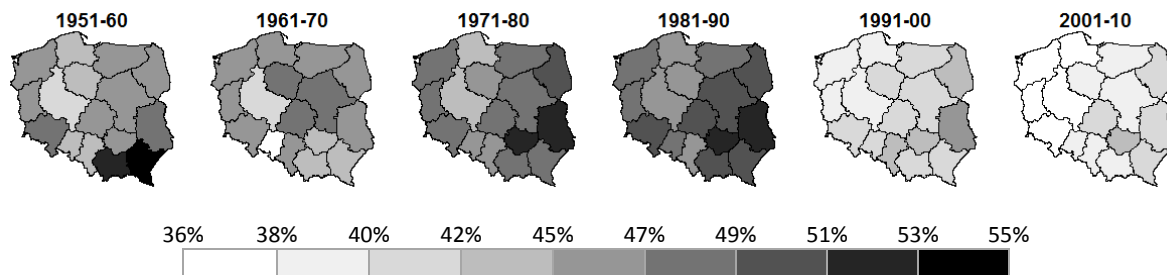
woj.	żeńskie						męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	50,3%	47,7%	52,4%	49,2%	39,9%	40,8%	46,9%	45,2%	47,9%	48,9%	40,5%	38,0%
KUJ.	50,6%	44,5%	50,5%	49,9%	39,2%	43,7%	42,7%	46,8%	46,3%	46,4%	41,4%	38,6%
LBL.	51,8%	49,1%	54,5%	51,3%	41,8%	42,9%	48,0%	45,5%	51,2%	51,5%	44,9%	41,8%
LBS.	49,9%	46,9%	51,5%	47,4%	38,6%	41,1%	45,2%	44,8%	47,5%	47,7%	39,6%	37,2%
ŁDZ.	50,2%	47,8%	51,8%	50,0%	41,3%	44,9%	45,6%	48,0%	47,8%	49,5%	41,9%	41,4%
MŁP.	52,2%	45,3%	48,0%	49,1%	42,3%	40,9%	51,9%	43,6%	47,2%	49,1%	41,1%	39,1%
MAZ.	48,9%	48,1%	54,4%	49,7%	41,7%	43,1%	45,0%	47,4%	48,4%	49,9%	42,0%	39,2%
OPL.	46,3%	38,8%	45,5%	45,7%	38,7%	40,1%	43,1%	37,8%	45,4%	48,5%	41,9%	39,1%
PDK.	55,2%	46,2%	48,1%	50,6%	40,7%	41,4%	53,7%	43,5%	48,1%	49,4%	41,5%	40,6%
PDL.	50,1%	48,0%	51,8%	51,4%	39,9%	43,2%	44,9%	45,3%	49,0%	49,7%	42,7%	41,9%
POM.	50,5%	41,7%	47,7%	47,9%	39,8%	39,8%	44,0%	45,0%	44,0%	45,6%	39,6%	36,0%
ŚLS.	45,2%	42,6%	46,9%	46,9%	40,3%	42,0%	43,3%	44,6%	46,3%	46,6%	43,4%	39,3%
ŚWK.	50,4%	48,8%	52,0%	51,6%	43,5%	45,6%	46,6%	44,2%	51,2%	51,5%	43,0%	42,7%
WRM.	52,3%	47,6%	51,9%	47,5%	40,7%	43,1%	45,9%	46,6%	48,5%	47,8%	41,3%	39,3%
WLP.	50,1%	42,7%	47,3%	48,9%	36,8%	40,6%	41,2%	42,0%	43,2%	45,1%	39,6%	37,9%
ZPM.	49,9%	46,7%	52,3%	48,7%	39,9%	43,2%	45,8%	46,3%	47,7%	47,8%	39,0%	37,9%
śred.	50,2%	45,8%	50,4%	49,1%	40,3%	42,3%	45,9%	44,8%	47,5%	48,4%	41,5%	39,4%
v	4,55	6,46	5,50	3,48	3,98	4,01	7,01	5,41	4,54	3,87	3,82	4,73

Rys. 56. Pokrycie sumy użyć przez 10 najczęstszych imion na obszarze współczesnych województw (dane z tab. 21)

imiona żeńskie



imiona męskie

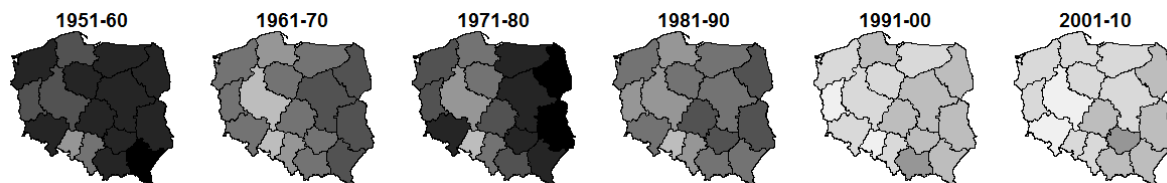


Tab. 22. Pokrycie sumy użyć przez 20 najczęstszych imion na obszarze współczesnych województw

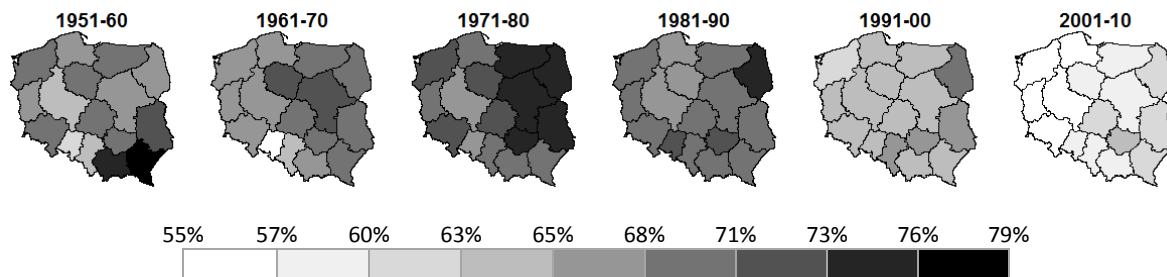
woj.	żeńskie						męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10	1951-60	1961-70	1971-80	1981-90	1991-00	2001-10
DLN.	74,0%	69,0%	74,0%	69,4%	60,7%	59,7%	69,8%	67,4%	71,1%	69,1%	63,9%	55,6%
KUJ.	73,8%	68,4%	69,5%	70,2%	62,2%	62,0%	69,4%	71,8%	71,9%	66,6%	64,2%	57,5%
LBL.	74,7%	71,9%	78,8%	72,3%	64,7%	64,3%	72,0%	69,7%	75,4%	70,5%	67,3%	61,7%
LBS.	72,8%	68,9%	71,9%	67,6%	59,3%	60,0%	67,8%	66,1%	70,6%	68,9%	62,9%	55,1%
ŁDZ.	76,0%	69,6%	73,2%	71,7%	63,8%	63,7%	68,7%	69,9%	72,0%	69,4%	65,0%	60,9%
MŁP.	75,5%	70,2%	71,8%	69,3%	65,4%	62,7%	74,2%	67,2%	69,4%	68,7%	63,9%	58,6%
MAZ.	74,4%	71,9%	75,6%	70,0%	63,0%	62,0%	66,0%	70,9%	74,2%	69,3%	64,3%	58,2%
OPL.	67,1%	63,6%	64,9%	65,1%	59,0%	60,3%	61,2%	57,2%	66,3%	70,8%	64,2%	59,4%
PDK.	76,2%	72,6%	73,4%	70,2%	63,8%	64,0%	76,4%	68,3%	70,0%	70,0%	64,8%	60,6%
PDL.	73,5%	71,1%	76,6%	73,3%	62,7%	64,6%	66,8%	68,9%	73,5%	73,7%	69,0%	61,8%
POM.	72,1%	66,7%	68,2%	67,5%	62,7%	60,2%	67,8%	67,8%	69,0%	66,2%	63,0%	54,5%
ŚLS.	70,2%	66,0%	68,9%	66,4%	61,8%	62,0%	64,5%	64,2%	70,4%	69,7%	66,4%	58,6%
ŚWK.	73,9%	70,2%	75,5%	72,2%	64,9%	66,8%	70,3%	68,5%	74,4%	72,2%	66,2%	63,1%
WRM.	74,6%	70,5%	73,5%	68,6%	61,0%	62,6%	69,9%	69,0%	73,9%	69,1%	64,3%	57,6%
WLP.	71,2%	64,5%	67,3%	67,1%	60,7%	59,2%	64,4%	65,9%	67,7%	66,2%	63,7%	57,0%
ZPM.	74,2%	68,4%	72,5%	69,3%	60,1%	61,8%	69,1%	67,8%	71,0%	68,2%	62,0%	55,7%
śred.	73,4%	69,0%	72,2%	69,4%	62,2%	62,2%	68,6%	67,5%	71,3%	69,3%	64,7%	58,5%
v	3,16	3,82	5,11	3,29	3,20	3,34	5,48	4,97	3,56	2,89	2,73	4,44

Rys. 57. Pokrycie sumy użyć przez 20 najczęstszych imion na obszarze współczesnych województw (dane z tab. 22)

imiona żeńskie



imiona męskie



C) Pokrycie sumy użyć przez 10 i 20 najczęstszych imion w imiennictwie wybranych miast na tle województw (lata 1951–2010)

Tab. 23. Porycie sumy użyć przez 10 najczęstszych imion w ośrodkach wielkomijskich na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	53%	53%	58%	53%	41%	40%	49%	50%	49%	51%	40%	37%
dśl.	50%	47%	51%	49%	40%	42%	48%	45%	48%	49%	41%	39%
różnica	3%	6%	8%	4%	1%	-2%	1%	6%	1%	2%	-1%	-2%
Kraków	55%	51%	57%	54%	44%	41%	50%	47%	52%	52%	42%	38%
młp.	54%	45%	46%	48%	42%	41%	54%	45%	46%	48%	41%	40%
różnica	1%	6%	11%	6%	3%	0%	-3%	2%	6%	4%	1%	-2%
Warszawa	55%	58%	58%	52%	41%	40%	47%	55%	51%	50%	40%	37%
maz.	50%	47%	54%	50%	43%	47%	46%	47%	49%	51%	44%	43%
różnica	5%	11%	4%	2%	-2%	-6%	1%	8%	1%	-2%	-3%	-6%
Poznań	53%	49%	56%	54%	38%	40%	46%	50%	49%	49%	40%	37%
włp.	51%	43%	45%	48%	37%	42%	42%	42%	43%	45%	40%	39%
różnica	2%	6%	11%	6%	1%	-2%	3%	8%	6%	5%	0%	-2%

Tab. 24. Porycie sumy użyć przez 20 najczęstszych imion w ośrodkach wielkomijskich na tle województw

	imiona żeńskie						imiona męskie					
	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10	1951-60	1961-70	1971-80	1981-90	1991-00	2000-10
Wrocław	76%	74%	77%	72%	62%	59%	70%	74%	73%	70%	63%	56%
dśl.	74%	69%	73%	69%	61%	61%	71%	67%	72%	69%	64%	57%
różnica	3%	5%	4%	4%	2%	-2%	-1%	6%	2%	1%	-1%	-1%
Kraków	77%	76%	78%	71%	66%	62%	74%	72%	76%	71%	65%	59%
młp.	76%	71%	70%	69%	65%	64%	76%	69%	69%	68%	64%	59%
różnica	1%	5%	8%	3%	1%	-2%	-2%	3%	7%	3%	1%	0%
Warszawa	75%	79%	76%	70%	63%	59%	69%	76%	74%	68%	62%	57%
maz.	76%	71%	77%	71%	65%	67%	69%	70%	76%	71%	67%	63%
różnica	-1%	8%	-1%	-1%	-3%	-8%	1%	6%	-2%	-3%	-5%	-6%
Poznań	74%	70%	73%	70%	63%	58%	67%	76%	71%	69%	64%	57%
włp.	73%	65%	66%	66%	61%	60%	66%	65%	67%	66%	64%	58%
różnica	2%	6%	6%	4%	2%	-2%	2%	10%	4%	4%	0%	-2%

Z danych zamieszczonych w tabeli 23 wynika, że we wszystkich miastach poziom badanego wskaźnika był zbliżony, wszędzie też w jego zakresie zaobserwowano tendencję opadającą. W pierwszych dekadach pokrycie sumy użyć przez 10 najczęstszych antroponimów było zwykle wyższe w miastach niż w województwach, natomiast w od czwartej dekady (w przypadku imiennictwa męskiego Warszawy od trzeciej) dochodzi do odwrócenia tego trendu lub przynajmniej znacznej niwelacji tych różnic. Te same spostrzeżenia dotyczą pokrycia sumy użyć przez 20 najczęstszych imion (tab. 24), z tym że wspomniany przed chwilą proces dokonał się w Warszawie jeszcze wcześniej, bo w 2–3 dekadzie. We wszystkich miastach odnotowano też znacznie większe różnice między imiennictwem żeńskim i męskim w zakresie badanych wskaźników (wyższa koncentracja

imion żeńskich) niż na pozostałym obszarze województw. Z czasem zmały zarówno różnice między imionami żeńskimi i męskimi w miastach, jak i poziom tych różnic między ośrodkami wielkomijskimi a prowincją.

III.4. Podsumowanie wyników analizy

Szczegółowa analiza poszczególnych parametrów rozkładu częstości pozwala na wysnuć pewnych wniosków dotyczących ogólnych tendencji z strukturze ilościowej imiennictwa w Polsce. Przykładowo w niemal każdym przypadku w imiennictwie ogólnopolskim dały się zaobserwować wyraźne trendy czasowe: zwiększanie⁶² się bogactwa leksykalnego do ok. połowy lat 40., następnie jego mniej lub bardziej gwałtowne zmniejszenie się po tym okresie i ponowny wzrost, zwykle od lat 80. Przeważnie też na końcu badanego okresu mieliśmy do czynienia z najwyższym poziomem bogactwa leksykalnego mierzonego poszczególnymi wskaźnikami niż na jego początku. Nie znaczy to jednak, że wszystkie przebadane ilościowe indykatory bogactwa leksykalnego pokazują to samo – mówimy bowiem tylko o pewnych zbieżnościach. Zanim więc przejdę do przedstawienia najbardziej charakterystycznych zjawisk, jakie ujawniły się w toku analizy, omówienia będzie wymagała sygnalizowana już wcześniej kwestia, mianowicie ich współwystępowania (bądź jego braku) w zakresie poszczególnych wskaźników. Z tabel 25 i 26 zawierających macierze korelacji liniowej Pearsona między wszystkimi omówionymi parametrami rozkładu częstości jasno wynika, że zmienność ich wartości w czasie była różna, czasem wręcz różniła się diametralnie.

Tab. 25. Macierz korelacji między wartościami parametrów rozkładu częstości w ogólnopolskim imiennictwie żeńskim w kolejnych latach z przedziału 1906–2010

	wielkość słownika	hapaksleg.	disleg.	trisleg.	oryginal.	K Yule'a	$I_{stereot}$	α	L	I_{konc}	pokr. 10	pokr. 20
w. słownika	1	0,98	0,98	0,97	0,74	-0,75	-0,95	-0,83	0,67	0,92	-0,75	-0,87
hapaksleg.	0,98	1	0,95	0,92	0,85	-0,70	-0,87	-0,72	0,76	0,83	-0,74	-0,84
disleg.	0,98	0,95	1	0,97	0,70	-0,72	-0,93	-0,81	0,67	0,89	-0,73	-0,82
trisleg.	0,97	0,92	0,97	1	0,63	-0,75	-0,95	-0,84	0,60	0,92	-0,73	-0,83
oryginal.	0,74	0,85	0,70	0,63	1	-0,43	-0,51	-0,31	0,89	0,44	-0,57	-0,59
K Yule'a	-0,75	-0,70	-0,72	-0,75	-0,43	1	0,79	0,68	-0,20	-0,78	0,96	0,88
$I_{stereot}$	-0,95	-0,87	-0,93	-0,95	-0,51	0,79	1	0,94	-0,46	-0,99	0,72	0,87
α	-0,83	-0,72	-0,81	-0,84	-0,31	0,68	0,94	1	-0,23	-0,96	0,57	0,80
L	0,67	0,76	0,67	0,60	0,89	-0,20	-0,46	-0,23	1	0,37	-0,32	-0,34
I_{konc}	0,92	0,83	0,89	0,92	0,44	-0,78	-0,99	-0,96	0,37	1	-0,69	-0,86
pokr. 10	-0,75	-0,74	-0,73	-0,73	-0,57	0,96	0,72	0,57	-0,32	-0,69	1	0,88
pokr. 20	-0,87	-0,84	-0,82	-0,83	-0,59	0,88	0,87	0,80	-0,34	-0,86	0,88	1

⁶² Nie zawsze tożsame ze wzrostem wartości danego wskaźnika – przykładowo przypomnijmy, że wyższe wartości wskaźnika koncentracji Mistrika wskazują na niższą koncentrację.

Tab. 26. Macierz korelacji między wartościami parametrów rozkładu częstości w ogólnopolskim imiennictwie męskim w kolejnych latach z przedziału 1906–2010

	słownik	hapaksleg.	disleg.	trisleg.	oryginal.	K Yule'a	$I_{stereot}$	α	L	I_{konc}	pokr. 10	pokr. 20
słownik	1	0,95	0,98	0,96	0,50	-0,32	-0,85	-0,76	0,62	0,74	-0,48	-0,73
hapaksleg.	0,95	1	0,95	0,87	0,74	-0,47	-0,65	-0,54	0,62	0,50	-0,61	-0,8
disleg.	0,98	0,95	1	0,95	0,56	-0,37	-0,80	-0,67	0,64	0,65	-0,53	-0,74
trisleg.	0,96	0,87	0,95	1	0,36	-0,23	-0,91	-0,78	0,62	0,79	-0,40	-0,65
oryginal.	0,50	0,74	0,56	0,36	1	-0,59	0,01	0,12	0,44	-0,20	-0,64	-0,61
K Yule'a	-0,32	-0,47	-0,37	-0,23	-0,59	1	-0,03	0,05	0,28	0,12	0,96	0,79
$I_{stereot}$	-0,85	-0,65	-0,8	-0,91	0,01	-0,03	1	0,90	-0,55	-0,96	0,13	0,43
α	-0,76	-0,54	-0,67	-0,78	0,12	0,05	0,90	1	-0,23	-0,95	0,18	0,46
L	0,62	0,62	0,64	0,62	0,44	0,28	-0,55	-0,23	1	0,38	0,14	-0,07
I_{konc}	0,74	0,50	0,65	0,79	-0,20	0,12	-0,96	-0,95	0,38	1	-0,01	-0,33
pokr. 10	-0,48	-0,61	-0,53	-0,40	-0,64	0,96	0,13	0,18	0,14	-0,01	1	0,89
pokr. 20	-0,73	-0,8	-0,74	-0,65	-0,61	0,79	0,43	0,46	-0,07	-0,33	0,89	1

Jak widzimy, pierwszy z badanych przez nas wskaźników, tj. wielkość słownika (W) w bardzo wysokim stopniu ($\rho > 0,9$) koreluje z liczbą hapaks-, dis-, trislegomenów, a także (w obrębie imiennictwa żeńskiego) ze wskaźnikami stereotypowości i koncentracji Mistríka. Generalnie wielkość słownika cechują stosunkowo wysokie wartości współczynnika korelacji z pozostałymi wskaźnikami. Najmniejsza zależność wystąpiła tu w zestawieniu z wskaźnikiem K Yule'a w obrębie antroponimii męskiej ($\rho = -0,32$, co odpowiada słabej korelacji ujemnej), choć w imiennictwie żeńskim zależność między tymi wskaźnikami była dużo wyższa ($\rho = -0,75$). Już samo to pokazuje, że należy z dużą ostrożnością wyciągać wnioski na temat bogactwa leksykalnego w antroponimii na podstawie tylko jednego ze wskaźników, a zwróćmy uwagę na fakt, że korelacja pomiędzy jeszcze innymi wskaźnikami była niekiedy niemal zerowa. Przykładowo w imiennictwie męskim nie występują żadne zależności (lub są marginalne) między wskaźnikiem stereotypowości Mistríka a wskaźnikami oryginalności Guirauda i różnorodności Yule'a czy między pokryciem sumy użyć przez 10 najczęstszych imion a wskaźnikiem koncentracji Mistríka. Co więcej, nawet jeśli między dwoma wskaźnikami zachodzi silna korelacja, to nie oznacza to, że wskaźniki te muszą być w zbliżonym do siebie stopniu skorelowane z innym wskaźnikiem. Tak jest np. w imiennictwie żeńskim w wypadku parametru α i wielkości słownika, dla których współczynnik korelacji wyniósł $-0,85$ (a zatem jest ona silna), jednak ich porównanie ze wskaźnikiem oryginalności wypada odmiennie (odpowiednio: $\rho = -0,31$ i $\rho = 0,74$).

Mając na względzie powyższe uwagi dotyczące zróżnicowania korelacji pomiędzy poszczególnymi wskaźnikami bogactwa leksykalnego, przejdźmy do przeglądu najważniejszych tendencji, które ujawniły się w toku analizy.

Wspomniałem już na początku o charakterystycznych cezurach wyznaczających zmianę trendu w zakresie poszczególnych wskaźników. Dotyczy to przede wszystkim pierwszej połowy lat 40., która niemal zawsze w mniejszym lub większym stopniu była momentem załamania dotychczasowego trendu świadczącego o wzroście bogactwa leksykalnego badanych zbiorów antroponimów. Największe zmiany dokonały się tu w zakresie wielkości słownika, udziału najradszej leksyki, stereotypowości, parametru α oraz wskaźnika koncentracji Mistríka. Jeśli chodzi o wskaźniki oryginalności Guirauda, K Yule'a, a także pokrycie sumy użyć przez 10 i 20 najczęstszych imion również notowano zmianę trendu, jednak jej przebieg był mniej gwałtowny, zaś w przypadku L Lorenza doszło do istotnej zmiany, lecz w kierunku odwrotnym – koncentracja po roku 1945 r. zaczęła się zmniejszać. W późniejszym okresie krzywe na poszczególnych wykresach mogły podlegać różnym fluktuacjom, jednak dla większości z nich wspólnym elementem jest okres 1980–85, gdy rozpoczynał się zwykle trend oddający wzrost bogactwa leksykalnego, który na przełomie lat 80. i 90. ulegał znacznej intensyfikacji kontynuowanej w późniejszych latach⁶³.

Przyczyny zaobserwowanych powyżej zjawisk mogą być różne, jednak wydaje się nie być kwestią przypadku ich pokrywanie się w czasie z momentami zwrotnymi w historii Polski, mającymi wpływ na przemiany społeczne, które mogły skutkować również przemianami w zachowaniach nazewniczych. Chodzi tu przede wszystkim o wyraźne różnice w trendach między okresem do 1945 r., okresem PRL oraz czasem transformacji ustrojowej przełomu lat 80. i 90. i kolejnymi latami. Można podejrzewać, że znaczny spadek różnorodności imiennictwa po II wojnie światowej był efektem szerszej zakrojonych, charakterystycznych dla PRL, dążeń do unifikacji kulturowej i społecznej. Odwrócenie tej tendencji w latach 80. i jej wzmocnienie w okresie III RP można z kolei wiązać ze znaczną pluralizacją życia społecznego w tym czasie⁶⁴.

Kolejnym często obserwowanym zjawiskiem była istotna różnica między imiennictwem żeńskim i męskim. Niemal przez cały rozpatrywany okres zasób imion żeńskich odznaczał się wyższym bogactwem opisywanym przy pomocy charakterystyk takich jak wielkość słownika, udział najradszej leksyki (zwłaszcza dis- i trislegomenów), wskaźnik

⁶³ Wyjątek stanowi tu wskaźnik oryginalności, którego regularny wzrost obserwujemy już od roku 1960, podobnie jak wskaźnik koncentracji Lorenza w obrębie antroponimii męskiej.

⁶⁴ Szerzej piszę o tym w zakończeniu (rozdział VI).

stereotypowości oraz wskaźnik koncentracji Mistríka. Przy innych miarach – K Yule’a, wskaźnikach koncentracji w ujęciu Lieberzona czy parametrze α – możemy mówić o trwałej przewadze tylko do połowy lat 40., później w krótszych interwałach szala przeważała się raz na korzyść imion męskich, raz żeńskich. W przypadku wskaźnika oryginalności Guirauda zmiana ta nastąpiła pod koniec lat 50., później mieliśmy do czynienia już z trwałą przewagą imiennictwa męskiego. Dała się jednak zauważyć również tendencja polegająca na niwelacji w czasie różnic między imiennictwem męskim i żeńskim. Stopniowe zmniejszanie od początku badanego okresu do jego końca obserwowaliśmy przy analizie wielkości słownika, najrzadszej leksyki, K Yule’a oraz L Lorenza. W zakresie parametru α oraz wskaźników stereotypowości i koncentracji Mistríka możemy jedynie stwierdzić, że różnice te zmalały znacznie w ostatniej dekadzie. Pozostałe wskaźniki wykazują w tej kwestii duże wahania.

Analiza parametrów rozkładu częstości wyrazów w zbiorach imion w poszczególnych województwach w latach 1951–2010 pokazała, że w Polsce istniało bądź istnieje geograficzne zróżnicowanie ilościowe imiennictwa. Ujawniły się pewne trwałe konfiguracje przestrzenne związane ze zbliżoną realizacją danej statystyki w sąsiadujących ze sobą regionach. Prześledzenie wartości wszystkich wskaźników w każdej z tabel oraz porównanie kartogramów może nastroić pewnych trudności. Na potrzeby niniejszego podsumowania, w celu uwypuklenia najbardziej charakterystycznych konfiguracji geograficznych w zakresie wszystkich badanych parametrów we wszystkich dekadach, postanowiłem zobrazować je pomocniczo w formie grafów (rys. 58, 59), które sporządziłem w sposób automatyczny przy pomocy podstawowych narzędzi z zakresu analizy sieciowej⁶⁵. Na rys. 58 przedstawione najbardziej charakterystyczne połączenia między województwami na bazie wyników obliczeń zamieszczonych w tabelach w poszczególnych podrozdziałach. Grafy z rys. 59 bazują na tych samych obliczeniach z wyjątkiem województwa mazowieckiego, dla którego przeprowadziłem ponowne obliczenia wszystkich wskaźników z wyłączeniem danych dla

⁶⁵ Grafy zostały utworzone przy pomocy funkcji *graph_from_adjacency_matrix* i *mst* z pakietu *igraph* (Csárdi, Nepusz 2006), działającego w środowisku R. Ich podstawą są macierze podobieństwa (odległości) między każdą parą województw obliczone na podstawie zestandaryzowanych wartości wszystkich badanych wskaźników we wszystkich dekadach (więcej o macierzach podobieństwa zob. w rozdziałach V.1 i V.2). Łącznie jest to 60 zmiennych dla każdego województwa – wartości 10 wskaźników notowanych w 6 dekadach. Na bazie macierzy podobieństwa funkcja *graph_from_adjacency_matrix* tworzy graf ukazujący relacje pomiędzy poszczególnymi obiektami (województwami) – im bliższe wartości danych wskaźników między parą województw, tym większa jest waga ich połączenia (krawędzi) i mniejsza odległość między nimi na grafie. Funkcja *mst* (minimum spanning tree, minimalne drzewo rozpinające) wydobywa z grafu najistotniejsze połączenia, które zostały przedstawione na rys. 58 i 59. Korzystając z wymienionych narzędzi, poza określeniem typu grafu (ważony, nieskierowany) użyłem ich automatycznych ustawień, pozostawiając programowi automatyczny dobór odpowiednich algorytmów.

Warszawy, której imiennictwo, jak mogliśmy zauważyć, znacząco różniło się pod kątem cech ilościowych od pozostałej części województwa.

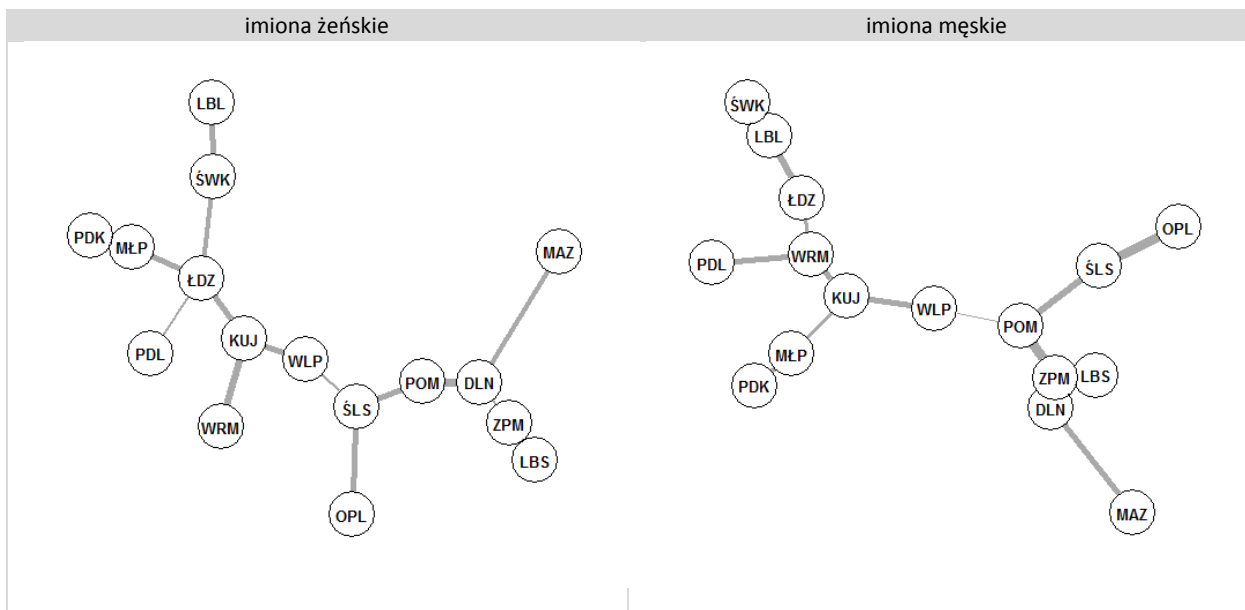
W toku analizy mogliśmy zaobserwować pewien generalny typ regionalizacji w zakresie realizacji poszczególnych parametrów rozkładu częstości w imiennictwie, mianowicie podział na część wschodnią i zachodnią Polski. Imiennictwo na zachodzie cechowało się na ogół większą różnorodnością niż na wschodzie. Oczywiście konfiguracja ta nie jest sztywna i widoczne były od niej pewne odstępstwa, jednak możemy mówić o swego rodzaju trendzie. Jednym z takich odstępstw było obserwowane w przypadku niemal wszystkich wskaźników ewidentne odchylenie województwa mazowieckiego nie tylko na tle sąsiadujących województw, ale całej Polski. Obszar ten charakteryzował się zdecydowanie bogatszym, bardziej różnorodnym imiennictwem. Jak wykazało badanie porównawcze ośrodków wielkomijskich na tle województw – wpływ na taki stan rzeczy miały przede wszystkim odstające wartości statystyk w samej Warszawie. Po wyłączeniu z obszaru woj. mazowieckiego danych dla stołecznego miasta okazało się, że wartości poszczególnych parametrów przypominają te dla sąsiadujących z nim województw lubelskiego i świętokrzyskiego, co możemy zobaczyć na rys. 59.

W ramach tego ogólnego podziału na wschód i zachód dały się zaobserwować również mniejsze zwarte obszary, charakteryzujące się podobną realizacją poszczególnych wskaźników, które w dużej mierze pokrywają się z określonymi regionami historyczno-kulturowymi. W województwie małopolskim i podkarpackim, pokrywających się w ogromnej mierze z obszarem dawnej Galicji, wartości wskaźników zwykle świadczyły o najniższym bogactwie leksykalnym imiennictwa. Niskie wartości notowano również w sąsiadujących z nimi od północy województwach lubelskim i świętokrzyskim (historycznie należącymi w większości do Małopolski), a także łódzkim i podlaskim które z kolei (wraz z woj. mazowieckim) należały do dawnego Królestwa Kongresowego. Na rys. 58 i 59 możemy zaobserwować, że woj. podlaskie odstaje nieco od pozostałych wschodnich województw – wiąże się to z faktem, iż zwłaszcza w ostatnich dekadach w zakresie niektórych wskaźników odznaczało się nieco bardziej różnorodnym imiennictwem.

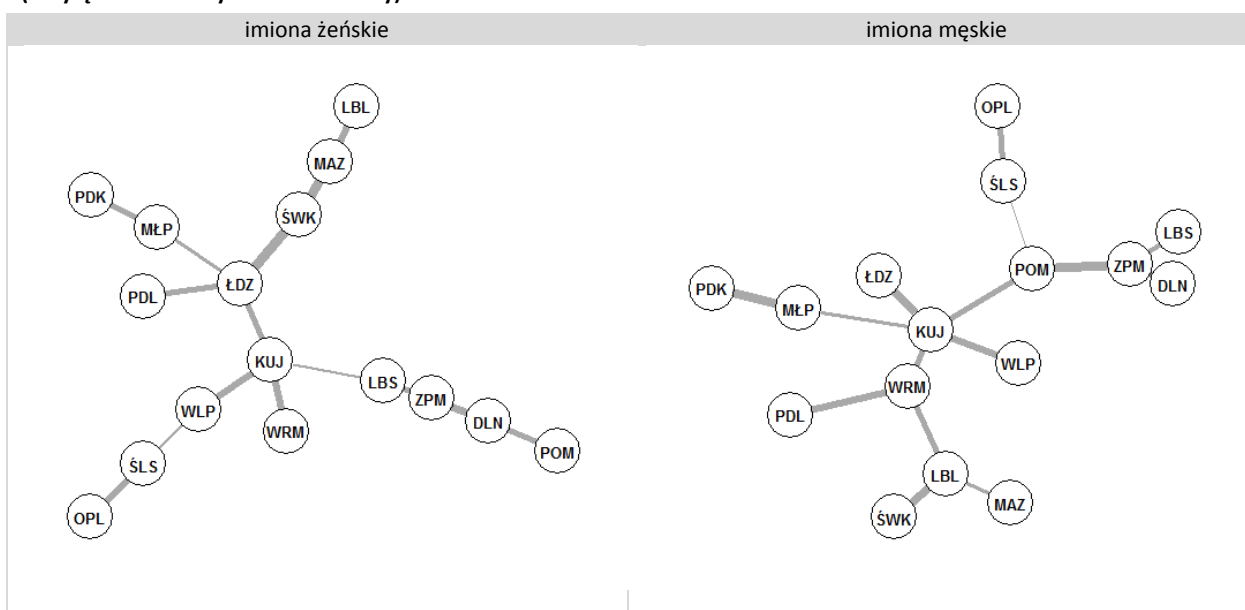
Na przeciwległym biegunie plasuje się pas województw wchodzących w obręb Ziemi Zachodnich, czyli dolnośląskie, lubuskie i zachodniopomorskie. Tu zdecydowanie najczęściej (nie licząc woj. mazowieckiego) w całym badanym okresie mieliśmy do czynienia z wartościami wskaźników świadczących o dużej różnorodności czy bogactwie zasobów imienniczych. Czasami łączyły się one w grupę również z woj. pomorskim (jego obszar

częściowo również przynależą do Ziemi Zachodnich). Równie charakterystyczne było dosyć częste współwystępowanie podobnych wartości niektórych wskaźników w sąsiadujących ze sobą woj. opolskim i śląskim, które w dużym stopniu pokrywają się z historycznym Górnym Śląskiem (woj. śląskie to obszar pograniczny śląsko-małopolski). Nierzadko też województwa te wykazywały podobieństwo do woj. wielkopolskiego i pomorskiego, z którymi razem należały w dużej mierze do dawnych Prus.

Rys. 58. Najbardziej charakterystyczne konfiguracje przestrzenne w zakresie badanych parametrów rozkładu częstości.



Rys. 59. Najbardziej charakterystyczne konfiguracje przestrzenne w zakresie badanych parametrów rozkładu częstości (z wyłączeniem danych dla Warszawy).



Sąsiadujące ze sobą województwa kujawsko-pomorskie i warmińsko-mazurskie na ogół wykazywały się raczej niższymi wartościami i bliżej im pod tym względem do obszaru wschodniego, co ujawniło się również na grafach z rys. 58 i 59, jednak łatwo zauważyć, że sytuują się one bliżej ich centralnej części, co w pewnym sensie koresponduje z ich pogranicznym położeniem w rzeczywistości. W pewnym stopniu można odnieść to również do woj. wielkopolskiego.

Jakkolwiek to zróżnicowanie geograficzne widoczne było na przestrzeni całego badanego okresu (1951–2010), to na podstawie analizy współczynnika zmienności (v), możemy stwierdzić, że w przypadku większości badanych wskaźników wraz z upływem czasu ulegało ono zacieraniu. Różnice między poszczególnymi województwami, zwłaszcza w ostatniej dekadzie (2001–2010), były zdecydowanie mniejsze niż w pierwszych dekadach powojennych. W ostatnim dziesięcioleciu widoczne były również coraz liczniejsze odstępstwa od obserwowanego wcześniej charakterystycznego układu przestrzennego danych.

Przy porównaniu pod kątem realizacji poszczególnych parametrów rozkładu częstości ośrodków wielkomiejskich z otaczającymi je terenami o niższym stopniu urbanizacji również zaobserwowane zostały pewne regularne zjawiska. Na ogół duże miasta charakteryzowały się dużo bardziej różnorodnym imiennictwem niż pozostały obszar województw, w których się znajdują – było to szczególnie widoczne, jak już wspomniano, na przykładzie Warszawy i woj. mazowieckiego. Jednak, podobnie jak w przypadku zróżnicowania między województwami, również i tu można było zwykle zaobserwować proces regularnej niwelacji różnic między miastami i prowincją, najwolniej jednak zachodzący w przypadku stolicy Polski. W miastach z reguły występowały też mniejsze niż na prowincji różnice w wartościach większości wskaźników między imionami żeńskimi i męskimi⁶⁶, szybciej też różnice te ulegały tam zmniejszeniu.

⁶⁶ Nie dotyczy to wskaźnika K Yule'a i pokrycia sumy użyć przez najczęstsze imiona, które wykazywały tendencję odwrotną.

IV. Tempo rotacji

W przywoływanych już w rozdziale II.1.3. pracach S. Liebersona i F. Lynn badacze, dysponując danymi dotyczącymi najpopularniejszych imion w Stanach Zjednoczonych oraz krajów Europy Zachodniej (również Węgier) zauważyli, że właściwie w każdym z nich na przestrzeni XX w. doszło do niespotykanego w poprzednich stuleciach przyspieszenia i wzrostu w zakresie wymiany najpopularniejszych imion w kolejnych sekwencjach czasu (Lieberson 2000: 36–66, Lieberson, Lynn 2003: 235, 255; Lieberson, Lynn 2008: 139–141). Do zobrazowania tego zjawiska autorzy posłużyli się prostym wskaźnikiem określanym jako tempo rotacji (ang. *turnover rate*), który opiera się na liczbie różnic pomiędzy listami frekwencyjnymi imion w poszczególnych punktach czasowych w obrębie imion o najniższych 10, 20 itd. rangach na listach frekwencyjnych.

Zasadniczo wskaźnikiem tempa rotacji może być po prostu liczba nowych pozycji wśród najczęstszych imion w stosunku do okresu poprzedzającego – przykładowo, jeśli lista dziesięciu najczęstszych imion różniła się między dwoma okresami trzema imionami, to tempo wymiany wynosi 30% (Doroszewicz 2013: 33, Gureckis, Goldstone 2009: 654). Jednak z uwagi na fakt, iż dla różnych krajów oraz warstw chronologicznych dysponowano danymi o różnej jakości, w przywołanych opracowaniach używa się wartości średniej rocznej liczby zmian, przy czym nie chodzi tu o zmiany na listach rangowych pomiędzy następującymi po sobie pojedynczymi latami, gdyż w tak krótkich odstępach czasu zmiany mogą być mało widoczne, ale w szerszych interwałach – 5-, 10-letnich i dłuższych (Lieberson 2000: 36, 290). W celu łatwiejszych porównań zestawów danych dla poszczególnych krajów stosuje się przeciętną liczbę zmian na rok (*turnover per annum*) – przykładowo jeśli w roku 1910 na liście dziesięciu najczęstszych imion pojawiły się dwa nowe w stosunku do listy z roku 1900, wówczas liczba ta dzielona jest przez dziesięć, co daje wynik 0,2 w skali roku.⁶⁷

W niniejszym rozdziale do zbadania dynamiki zmian w imiennictwie polskim posłużę się prostym wskaźnikiem tempa rotacji przedstawionym powyżej, który ujmuje liczbę różnic pomiędzy listami r najczęstszych imion w równych przedziałach czasowych (5- i 10-letnich) podzieloną przez liczbę lat z danego przedziału. Jest to podyktowane nie tylko łatwością

⁶⁷ S. Wilson, który badał tempo rotacji w imiennictwie różnych obszarów na przestrzeni wielu stuleci, dla starszych okresów często nie dysponował danymi, które pozwalałyby zmierzyć tempo rotacji w równych odstępach. Chcąc porównywać tempo zmian zmierzone w różnych odstępach, autor dokonywał interpolacji z okresu dłuższego na docelowy, np. dziesięcioletni. Jeśli zatem w ciągu 25 lat doszło do zmiany 4 imion na liście określonej liczby najczęstszych imion, to w skali dekady u Wilsona będzie to zmiana na poziomie 1,6 imienia (Wilson 1998: 326).

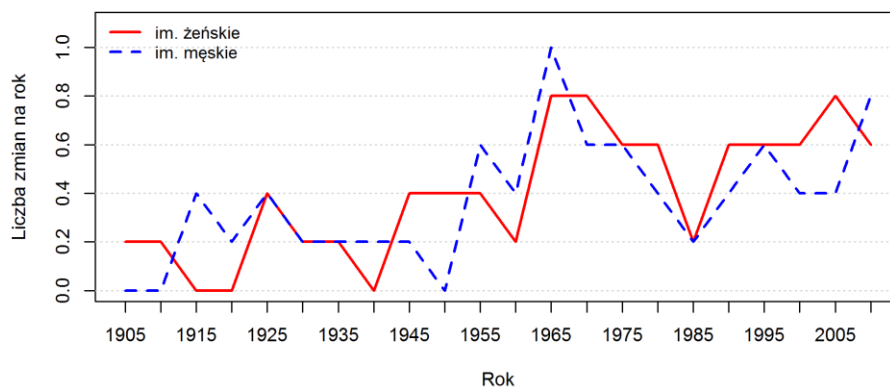
zastosowania tej metody, ale także możliwością porównania uzyskanych wyników z danymi dla wielu innych krajów, które przedstawili wspomniani wcześniej badacze. Zwróć jednak też uwagę na pewne wady tego sposobu analizy, gdyż jak się okazuje, dobór liczby najczęstszych imion oraz długość przedziałów czasowych między poszczególnymi punktami pomiarowymi wpływa na kształt zaobserwowanych zmian.

Ponadto w podrozdziale IV.4 analizie poddano także rotację w skali rocznej (czyli zmiany w zasobie r najczęstszych imion w danym roku w odniesieniu do roku poprzedzającego), w której ujawniły się zmiany nieuchwytnie przy badaniu zjawiska rotacji przy użyciu szerokich, sztywnych przedziałów czasowych.

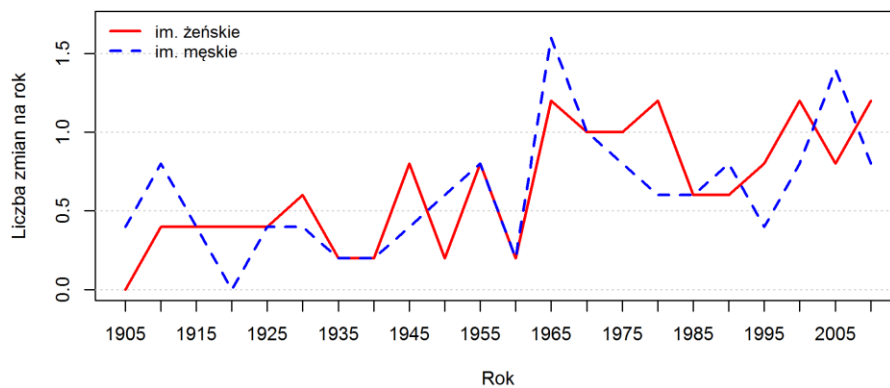
IV.1 Tempo rotacji w imiennictwie ogólnopolskim

Do zbadania tempa rotacji w ogólnopolskim imiennictwie wykorzystano listy frekwencyjne (osobno dla imion żeńskich i męskich) sporządzone dla każdego roku w przedziale czasowym 1900–2010. Punkty pomiarowe wyznaczone zostały na podstawie pięcio- i dziesięcioletnich przedziałów czasowych. W przypadku przedziałów pięcioletnich pierwszym punktem pomiarowym był rok 1905, natomiast przy przedziałach dziesięcioletnich – rok 1910. Ostatnim punktem pomiarowym w obu wypadkach był rok 2010. Badanie tempa wymiany przeprowadzono kolejno na danych dla pierwszych 10, 20, 30, 50 i 100 najczęstszych imion z list frekwencyjnych dla danego roku. Wyniki badania zostały przedstawione na serii wykresów (rys. 60–64 dla przedziałów 5-letnich i rys. 65–69 dla przedziałów 10-letnich). Na każdym z wykresów na osi x widnieją kolejne lata stanowiące punkty pomiarowe, natomiast na osi y średnia liczba zmian w obrębie r najczęstszych imion (będąca wynikiem dzielenia liczby nowych pozycji na liście najczęstszych imion w danym roku przez liczbę lat w przyjętym przedziale czasowym).

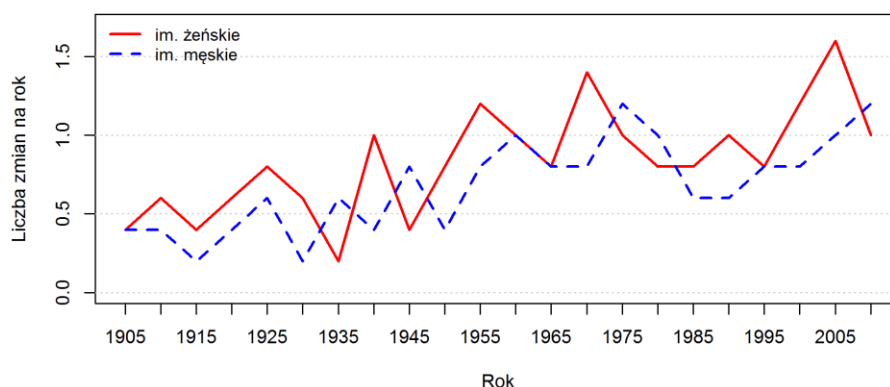
Rys. 60. Zmiany wśród 10 najczęstszych imion (odstępny 5-letnie)



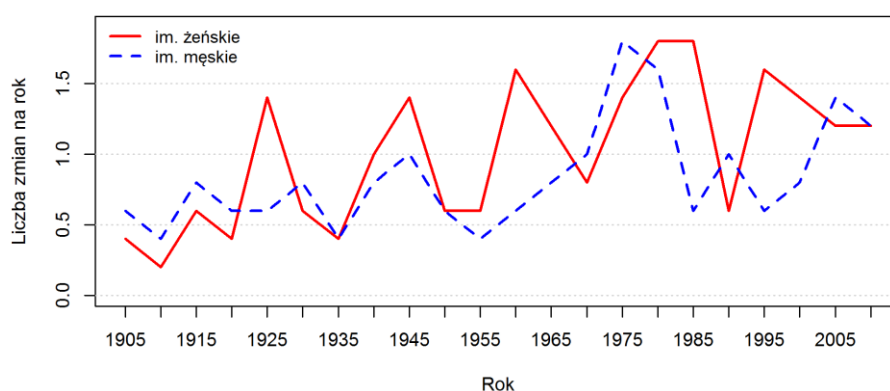
Rys. 61. Zmiany wśród 20 najczęstszych imion (odstępny 5-letnie)



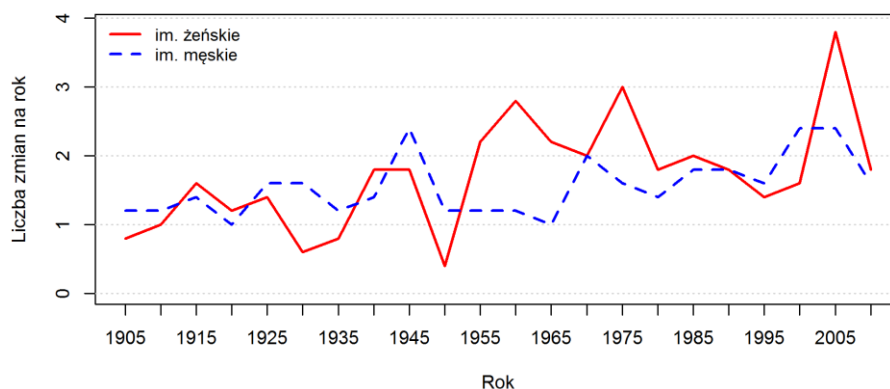
Rys. 62. Zmiany wśród 30 najczęstszych imion (odstęp 5-letnie)



Rys. 63. Zmiany wśród 50 najczęstszych imion (odstęp 5-letnie)



Rys. 64. Zmiany wśród 100 najczęstszych imion (odstęp 5-letnie)



Z przedstawionych na rys. 60–64 wykresów dla wskaźnika rotacji w obrębie 10, 20, 30, 50 i 100 najczęstszych imion (badanego w pięcioletnich odstępach) wynika, że w ciągu całego rozpatrywanego okresu liczba zmian w obrębie najczęstszych imion na listach frekwencyjnych wykazuje tendencję rosnącą, choć nie możemy mówić o stałym, regularnym wzroście, ale o trendzie. Widzimy bowiem, że niemal zawsze po wzroście liczby zmian następuje ich spadek, rzadko natomiast występuje zjawisko stałego wzrostu lub spadku (to samo zachowanie obserwujemy w przypadku wykresów Gureckisa i Goldstone’a oraz Hahna

i Bentleya, zob. rys. 12 i 13 w rozdziale II.1.3). Niemniej jednak w drugiej połowie badanego okresu wartości wskaźnika rotacji są niemal zawsze wyższe od tych w pierwszej połowie. Nasze obserwacje potwierdzają więc tezę S. Lieberzona oraz pozostałych badaczy o rosnącym tempie rotacji w XX wieku w obrębie najczęstszych imion, opartą na danych z innych krajów.

Uzyskane przeze mnie wartości dotyczące tempa rotacji są porównywalne z danymi dla innych krajów. W przypadku nowych pozycji na liście 10 najczęstszych imion we Francji i Niemczech w pierwszym kwartale XX w. wskaźnik rotacji wynosił między 0,2 a 0,5, natomiast od lat 50. utrzymywał się stale w przedziale między 0,5 do 1 (we Francji wyjątkowo osiągając wartości 1,4 czy 2,4) (Lieberson 2000: 37–38). Z kolei wyniki dla pierwszych 20 imion pokrywają się z tymi dostępnymi dla Danii (0,4–0,6 w pierwszej połowie XX w. i 0,6–1,1 w drugiej) i częściowo tylko dla Kalifornii (w przypadku imion męskich wskaźnik rotacji wynosił tu 0,0–0,5 w 1. poł. XX w. i powyżej 0,6 w drugiej; imiona żeńskie odznaczały się dużo wyższą zmiennością – odpowiednio 0,4–1,0 i 1,0–1,3) (tamże: 40).

Rotację w obrębie 50 najczęstszych imion możemy porównać z danymi dla Anglii i Walii – wartość wskaźnika w pierwszym kwartale XX w. wynosiła dla imion męskich 0,6–0,9, dla żeńskich – 0,9–1,6, zaś w drugiej połowie stulecia odpowiednio 0,8–1,8 i 1,6–1,8 (tamże: 39). Już pobieżna obserwacja powyższych wykresów pozwala zaobserwować bardzo podobne wartości, choć warto podkreślić, że należy zachować ostrożność przy tego typu porównaniach, gdyż wartości wskaźnika tempa rotacji mogą być zależne od długości przedziałów czasowych, w jakich był on mierzony. Do kwestii tej wrócę w dalszej części tego rozdziału.

O ile w zakresie wzrostu tempa rotacji ogólne tendencje w Polsce pokrywają się z tendencjami w innych krajach, to jednak wyniki naszych badań są nieco odmienne od tych uzyskanych przez Lieberzona, a także przez Hahna i Bentleya oraz Gureckisa i Goldstone'a, jeśli chodzi o różnice w tempie wymiany między imionami żeńskimi i męskimi. Podczas gdy u wspomnianych badaczy niemal zawsze wskaźnik rotacji był wyższy w przypadku imion żeńskich, to w imiennictwie polskim różnica ta nie jest tak wyraźna. W niektórych punktach liczba nowych imion męskich przewyższa liczbę zmian w obrębie imion żeńskich, w innych zaś jest odwrotnie. Jedynie na wykresach dla pierwszych 30 i 50 imion (rys. 62 i 63) można dopatrzeć się wyraźnie wyższych wartości dla imion żeńskich. W odróżnieniu od badanych w poprzednim rozdziale parametrów rozkładu częstości wizualna ocena różnic w zakresie tempa rotacji między imiennictwem żeńskim i męskim nastrocza pewnych trudności.

Odwołując się jednak do pracy Hahna i Bentleya (2003: 121), w celu określenia poziomu tych różnic możemy posłużyć się testem *t*-Studenta na istotność różnic dla zmiennych powiązanych, który pozwoli nam określić, czy różnica średnich wartości tempa rotacji w całym badanym okresie między imionami żeńskimi i męskimi była istotna statystycznie. W tab. 27 zamieszczono średnie wartości wskaźników dla całego rozpatrywanego okresu oraz wartości *p* z testu *t*-Studenta dla różnic między nimi⁶⁸.

Tab. 27. Wynik testu istotności różnic dla wskaźników rotacji w obrębie imion męskich i żeńskich (przedziały 5-letnie, 1905–2010)

<i>r</i> najcz. imion	średnia wartość wskaźnika rotacji		wartość <i>p</i>
	im. żeńskie	im. męskie	
10	0,40	0,37	0,561
20	0,65	0,62	0,701
30	0,84	0,68	0,029
50	1,01	0,84	0,106
100	1,72	1,55	0,319

Jak wynika z danych zamieszczonych w powyższej tabeli średnie tempo rotacji w każdym przypadku było zawsze wyższe w przypadku imion żeńskich. Jednak wzięwszy pod uwagę dane z wszystkich punktów czasowych, które składają się na tę średnią, jedynie w przypadku porównania wskaźnika rotacji w obrębie 30 najczęstszych imion możemy mówić o statystycznie istotnej różnicy (na poziomie $p < 0,05$) między imionami męskimi i żeńskimi. Oczywiście nie oznacza to, że różnice te nie występują w ogóle. Obserwując linie na poszczególnych wykresach, możemy zaobserwować, że w drugiej połowie badanego okresu osobom płci żeńskiej zdecydowanie częściej nadawano nowe imiona. Warto jednak zwrócić uwagę, że w zależności od doboru liczby najczęstszych imion możemy uzyskać bardzo różne wyniki i należy ostrożnie wyciągać wnioski z takich zestawień.

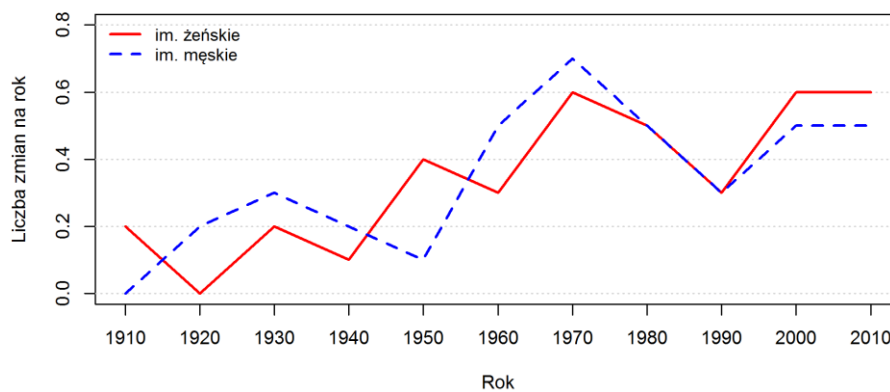
Opisana powyżej nieścisłość nie jest jedyną, która wynika z decyzji o arbitralnym doborze liczby najczęstszych imion do analizy tempa rotacji. Gdybyśmy bowiem szczegółowo chcieli przeanalizować zmiany, jakie zachodziły pomiędzy zbiorami imion w wyznaczonych pięcioletnimi interwałami punktach, to okaże się, że w zależności od liczby analizowanych imion możemy uzyskać sprzeczne wyniki. Przykładowo, w 2010 r. odnotowano wzrost liczby zmian w obrębie 10 najczęstszych imion męskich w stosunku do roku 2005 (4 imiona, średnio

⁶⁸ Zazwyczaj przyjmuje się, że różnice między średnimi są istotne, gdy $p < 0,05$ (istotność na poziomie 95%) lub $p < 0,1$ (istotność na poziomie 90%).

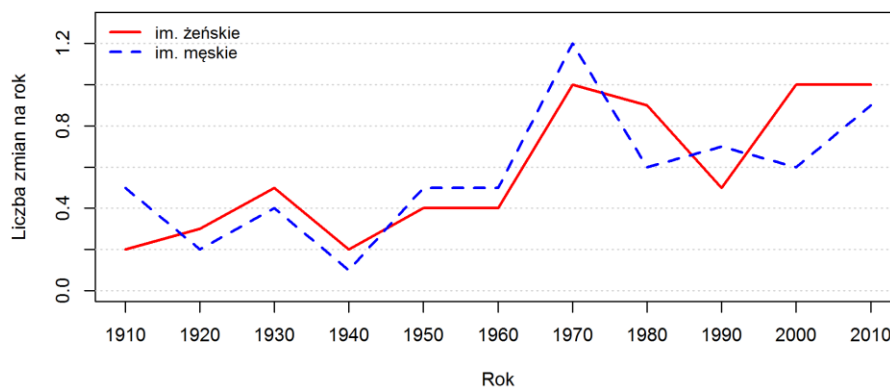
0,8 w skali roku), jednak w przypadku 20 najczęstszych imion taka sama różnica na liście oznaczała już spadek liczby zmian, ponieważ w obrębie tych dwudziestu imion nie zaszły żadne dodatkowe zmiany. Podobnych przykładów na przedstawionych wykresach możemy zaobserwować więcej.

Ostrożność należy zachować także przy stosowaniu różnej długości przedziałów czasowych między punktami pomiarowymi. Przyjrzyjmy się serii wykresów (rys. 65–69) dla wskaźnika tempa rotacji w obrębie najczęstszych 10, 20, 30, 50 i 100 imion, tym razem mierzonego w odstępach 10-letnich (w okresie 1910–2010):

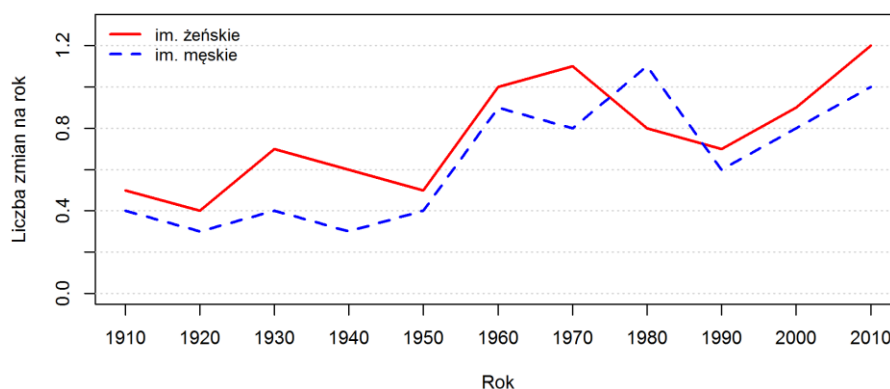
Rys. 65. Zmiany wśród 10 najczęstszych imion (odstępów 10-letnie)



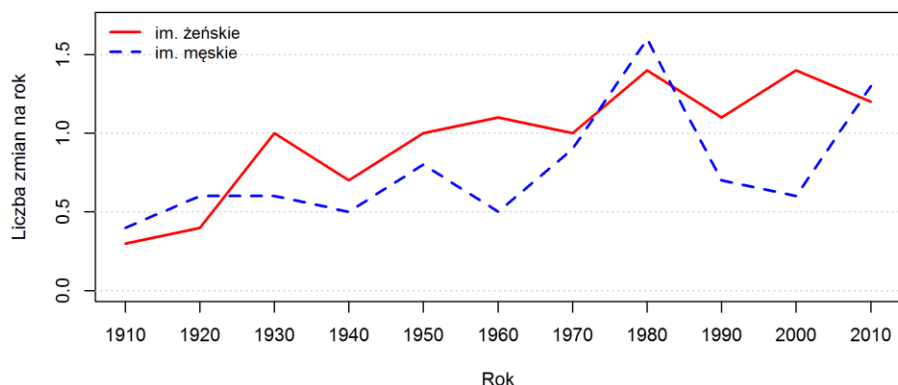
Rys. 66. Zmiany wśród 20 najczęstszych imion (odstępów 10-letnie)



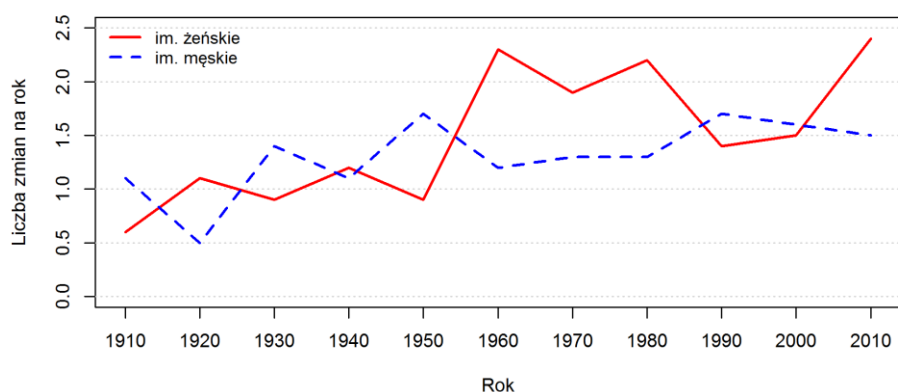
Rys. 67. Zmiany wśród 30 najczęstszych imion (odstępów 10-letnie)



Rys. 68. Zmiany wśród 50 najczęstszych imion (odstępy 10-letnie)



Rys. 69. Zmiany wśród 100 najczęstszych imion (odstępy 10-letnie)



Powyższe wykresy potwierdzają tendencję wzrostową, jeśli chodzi o liczbę zmian w obrębie najczęstszych imion. Podobnie też, jak to odnotowałem wcześniej, istotna przewaga wskaźnika rotacji w imiennictwie żeńskim daje się zauważyć jedynie w obrębie imion o najniższych 30 ($p < 0,05$) i 50 ($p < 0,1$) rangach na listach frekwencyjnych (tab. 28).

Tab. 28. Wynik testu istotności różnic dla wskaźników rotacji w obrębie imion męskich i żeńskich (przedziały 10-letnie, 1910–2010)

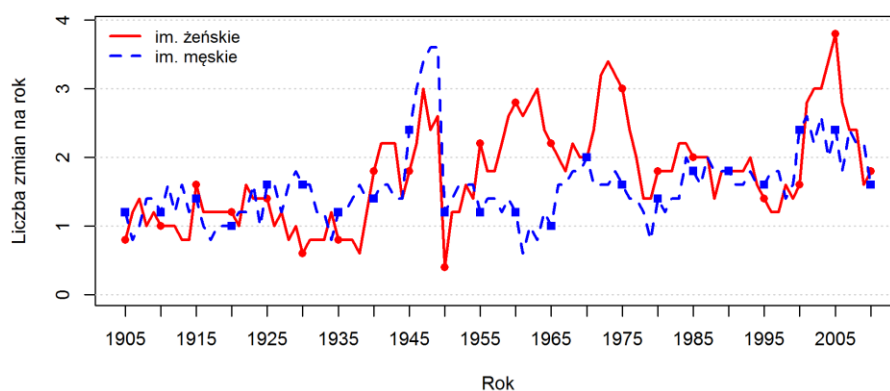
<i>r</i> najcz. imion	średnia wartość wskaźnika rotacji		wartość <i>p</i>
	im. żeńskie	im. męskie	
10	0,35	0,35	1,000
20	0,58	0,56	0,788
30	0,76	0,64	0,031
50	0,96	0,77	0,086
100	1,49	1,31	0,387

Łatwo jednak zauważyć, że wartości wskaźnika rotacji w skali rocznej, a także kierunek zmian w poszczególnych punktach pomiarowych nierzadko różni się od wykresów dla przedziałów pięcioletnich. Przykładowo zwróćmy uwagę, że na wykresach dotyczących rotacji w obrębie 100 najczęstszych imion dla przedziałów 5-letnich (rys. 64) otrzymaliśmy

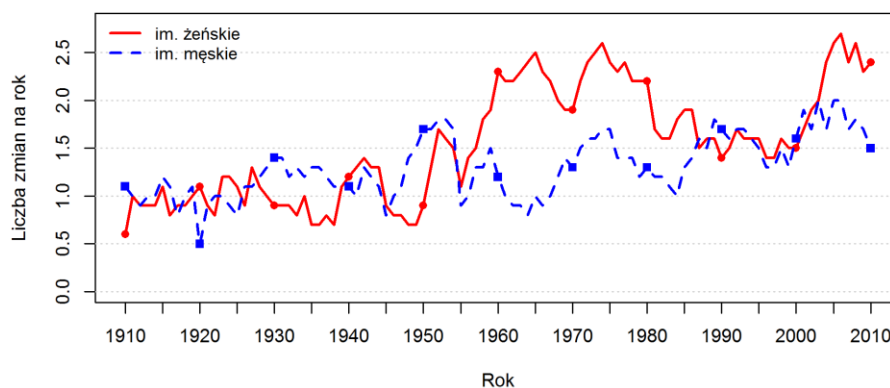
bardzo wysokie wartości w skali rocznej (maks. 3,8), natomiast dla przedziałów 10-letnich (rys. 69) były one dużo niższe (maks. 2,4). Pokazuje to, że interpolacja wskaźnika tempa rotacji, dokonywana przez Wilsona (1998: 326), niekoniecznie musi oddawać rzeczywiste tendencje. Choć oczywiście należy pamiętać, że w przypadku fragmentarycznych danych (zwłaszcza dla wcześniejszych stuleci), jakimi dysponował badacz zabieg taki jest rzeczywiście jedynym wyjściem.

Również mierzenie wskaźnika tempa rotacji w arbitralnie przyjętych punktach (rozdzielonych równymi odstępami), jak to czynią Hahn i Bentley oraz Gureckis i Goldstone, choć pozwala uchwycić pewne ogólne tendencje, a także jest wygodne i przejrzyste w wizualizacji, to jednak niesie ze sobą ryzyko przeoczenia pewnych wyraźnych zjawisk. Przyjrzyjmy się wykresom na rys. 70 i 71. Przedstawiono na nich, podobnie jak w przypadku rys. 64 i 69, zmiany w obrębie stu najczęstszych imion w stosunku do 5 i 10 lat wstecz, jednak tym razem nie tylko dla wybranych lat rozdzielonych równymi przedziałami czasowymi, ale dla każdego roku z przedziału 1905–2010 (dla odstępów 5-letnich) i 1910–2010 (dla odstępów 10-letnich).

Rys. 70. Zmiany wśród 100 najczęstszych imion dla każdego roku w stosunku do 5 lat wstecz



Rys. 71. Zmiany wśród 100 najczęstszych imion dla każdego roku w stosunku do 10 lat wstecz



Na powyższych wykresach (rys. 70 i 71) czerwonymi kołami i niebieskimi kwadratami dodatkowo zaznaczono punkty, przez które przebiegały linie na wykresach zamieszczonych na rys. 64 i 69. Obserwując górny wykres, łatwo zauważyć, że mierząc poziom wskaźnika rotacji tylko w wybranych latach (w 5-letnich sekwencjach), pominęliśmy na nich bardzo istotne zmiany w obrębie zarówno imion męskich, jak i żeńskich, jakie dokonywały się w latach 40. XX wieku. Dolny wykres pokazuje z kolei, że przy arbitralnie ustalonych 10-letnich sekwencjach dla lat 1910-2010, nie zostały uwzględnione maksymalne wartości, jakie wskaźnik rotacji osiągał w przypadku imion żeńskich w latach 60. i 70. oraz na początku XX w. Gdybyśmy sporządzając wcześniejsze wykresy za punkt startowy przyjęli daty inne niż 1905 czy 1910 r., wówczas linie przebiegałyby na nich w nieco odmienny sposób. Z kolei porównanie obu powyższych wykresów unaocznia nam, że badając zmiany w dłuższych odstępach czasowych, możemy w ogóle nie zauważyć zjawisk, jakie ujawniły się w przypadku krótszych odstępów (jak we wspomnianych już latach 40.). Wydaje się zatem uzasadnione badanie wskaźnika tempa rotacji w możliwie wielu konfiguracjach jeśli chodzi o liczbę najczęstszych imion oraz długość odstępów dzielących punkty pomiarowe (jak się okaże w podrozdziale IV.5, również badanie zmian pomiędzy następującymi latami, może dać interesujące wyniki).

Podsumowując tę część rozdziału, powtórzmy najistotniejsze wyniki naszych badań. Przeprowadzona analiza wskaźnika tempa rotacji w obrębie najczęstszych 10, 20, 30, 50 i 100 imion męskich i żeńskich w poszczególnych latach w odstępach 5- i 10-letnich potwierdziła, że w imiennictwie polskim doszło do wzrostu i przyspieszenia tempa rotacji na przestrzeni XX wieku. Może to świadczyć o tym, że podobnie jak w wielu innych krajach Europy oraz Ameryki Północnej, również w Polsce w coraz większym stopniu zachowania nazewnicze podlegają mechanizmom mody, a zwyczaj nazewniczy ustępuje miejsca indywidualnym gustom rodziców (Lieberson 2000: 66–68). Natomiast tylko częściowo wyniki analizy pokryły się z obserwacjami innych badaczy, dotyczącymi znacznie większego tempa rotacji w przypadku nadawania imion osobom płci żeńskiej. Trudno jest wskazać, jakie mogły być tego przyczyny, zwłaszcza, że w badaniu parametrów rozkładu częstości w poprzednim rozdziale różnice w zachowaniach nazewniczych ze względu na płeć nazywanych osób były często bardzo wyraźne.

IV.2. Tempo rotacji w imiennictwie na obszarze współczesnych województw

Porównanie tempa rotacji na obszarze województw ograniczyłem do prześledzenia liczby zmian w obrębie 10 najczęstszych imion w 5-letnich odstępach w okresie 1945–2010. Pierwszym punktem pomiarowym był rok 1950, ostatnim zaś 2010. Z uwagi na dużą ilość danych dotyczących wartości wskaźnika tempa rotacji ich prezentacja w postaci wykresów liniowych (jak powyżej) byłaby mało czytelna, dlatego – analogicznie do badania w zakresie parametrów rozkładu częstości – przedstawiam je w formie tabelarycznej (tab. 29). Dane z tabeli zostały także zwizualizowane w formie kartogramów (rys. 72), co ułatwi obserwacje geograficznego zróżnicowania poziomu tempa rotacji w poszczególnych punktach czasowych. Z kolei w tab. 30 porównano średnie wartości omawianego wskaźnika w imiennictwie żeńskim i męskim w poszczególnych województwach oraz wyniki testu t dla tych średnich.

Analiza danych zamieszczonych w tab. 29, naniesionych na mapy (rys. 72), odwrotnie niż w przypadku informacji o wskaźnikach bogactwa leksykalnego, nie ujawniła, by w zakresie tempa rotacji w poszczególnych latach występowała charakterystyczna i stała konfiguracja regionalna. W zależności od przedziału czasowego układ województw pod względem wartości wskaźnika tempa rotacji jest różny. Również gdy wyciągniemy średnią ze wszystkich lat (zob. tab. 30 i rys. 73) dla każdego z województw nie otrzymamy żadnych zwartych regionów. Średnio najwyższe tempo rotacji w imiennictwie żeńskim notowano w woj. opolskim (0,62), lubuskim, lubelskim (0,6), zachodniopomorskim, świętokrzyskim, podkarpackim i dolnośląskim (0,58), zaś w imiennictwie męskim w woj. lubuskim, kujawsko-pomorskim (0,55), zachodniopomorskim, wielkopolskim (0,54), lubelskim, świętokrzyskim i warmińsko-mazurskim (0,51).

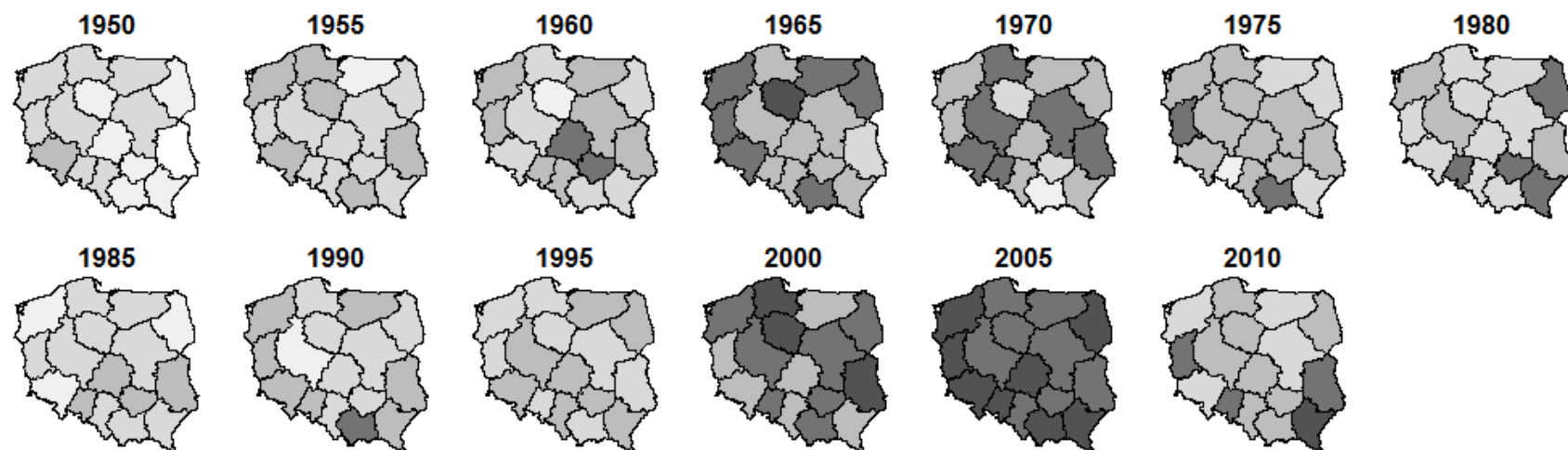
Na podstawie danych z tab. 30 i rys. 73 łatwo zauważyć, że we wszystkich województwach z wyjątkiem kujawsko-pomorskiego, tak jak w imiennictwie ogólnopolskim, mieliśmy do czynienia z większą średnią wartością wskaźnika rotacji w obrębie imiennictwa żeńskiego niż męskiego. Największe różnice wystąpiły w Polsce południowej, na obszarze województw dolnośląskiego, opolskiego, małopolskiego i śląskiego, jednak biorąc pod uwagę dane z wszystkich punktów czasowych o istotnie statystycznej różnicy (na poziomie $p < 0,1$) możemy mówić jedynie w przypadku województwa dolnośląskiego. Najmniejsze różnice między imiennictwem żeńskim i męskim odnotowano w województwach kujawsko-pomorskim, warmińsko-mazurskim i wielkopolskim.

Tab. 29. Liczba zmian na rok wśród 10 najczęstszych imion na obszarze współczesnych województw (odstępy 5-letnie)

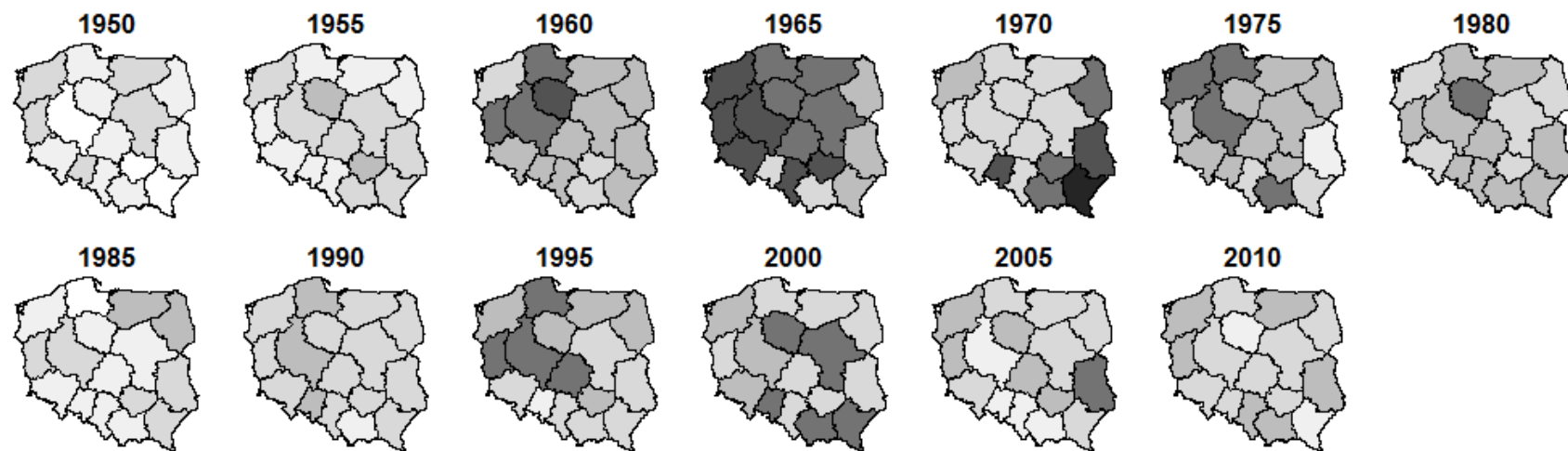
woj.	imiona żeńskie													imiona męskie												
	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010
DLN	0,6	0,6	0,4	0,8	0,8	0,6	0,4	0,2	0,6	0,6	0,6	1,0	0,4	0,2	0,2	0,6	1	0,4	0,6	0,4	0,2	0,4	0,4	0,6	0,4	0,4
KUJ	0,2	0,6	0,2	1,0	0,4	0,6	0,4	0,4	0,4	0,4	1,0	0,8	0,6	0,2	0,6	1	0,8	0,4	0,6	0,8	0,2	0,4	0,6	0,8	0,6	0,2
LBL	0,0	0,6	0,6	0,4	0,8	0,6	0,6	0,6	0,6	0,4	1,0	0,8	0,8	0,2	0,4	0,6	0,6	1	0,2	0,6	0,4	0,4	0,4	0,4	0,8	0,6
LBS	0,4	0,4	0,6	0,8	0,6	0,8	0,4	0,4	0,6	0,4	0,6	1,0	0,8	0,4	0,2	0,8	1	0,4	0,6	0,6	0,4	0,4	0,8	0,4	0,6	0,6
ŁDZ	0,2	0,4	0,8	0,6	0,6	0,6	0,4	0,6	0,4	0,6	0,6	1,0	0,6	0,2	0,4	0,6	0,8	0,4	0,6	0,6	0,2	0,4	0,8	0,4	0,6	0,4
MŁP	0,2	0,6	0,4	0,8	0,2	0,8	0,4	0,4	0,8	0,4	0,8	1,0	0,6	0,2	0,4	0,4	0,4	0,8	0,8	0,6	0,2	0,2	0,4	0,8	0,2	0,6
MAZ	0,4	0,4	0,6	0,6	0,8	0,6	0,4	0,4	0,4	0,4	0,8	0,8	0,4	0,4	0,4	0,6	0,8	0,4	0,6	0,4	0,2	0,4	0,4	0,8	0,4	0,4
OPL	0,4	0,4	0,6	0,6	0,8	0,2	0,8	0,6	0,6	0,4	0,8	1,0	0,8	0,4	0,2	0,6	0,4	1	0,6	0,6	0,2	0,6	0,2	0,8	0,2	0,4
PDK	0,2	0,4	0,4	0,6	0,6	0,4	0,8	0,4	0,6	0,6	0,6	1,0	1,0	0	0,4	0,6	0,6	1,2	0,4	0,6	0,4	0,4	0,4	0,8	0,4	0,2
PDL	0,2	0,4	0,4	0,8	0,6	0,4	0,8	0,2	0,4	0,6	0,8	1,0	0,6	0,2	0,2	0,6	0,6	0,8	0,6	0,4	0,6	0,4	0,6	0,4	0,4	0,4
POM	0,4	0,6	0,4	0,6	0,8	0,6	0,4	0,4	0,4	0,4	1,0	0,8	0,6	0,2	0,2	0,8	0,8	0,4	0,8	0,6	0	0,6	0,8	0,4	0,4	0,4
ŚLS	0,4	0,4	0,6	0,6	0,6	0,6	0,4	0,4	0,4	0,6	0,6	0,8	0,6	0,2	0,2	0,6	1	0,4	0,4	0,6	0,2	0,4	0,4	0,4	0,2	0,6
ŚWK	0,2	0,4	0,8	0,6	0,4	0,6	0,8	0,6	0,4	0,6	0,8	0,8	0,6	0	0,6	0,4	1	0,8	0,6	0,4	0,4	0,4	0,6	0,4	0,6	0,4
WRM	0,4	0,2	0,6	0,8	0,6	0,4	0,4	0,4	0,6	0,6	0,6	0,8	0,4	0,4	0,2	0,6	0,8	0,4	0,6	0,6	0,6	0,4	0,6	0,4	0,4	0,6
WLP	0,4	0,4	0,4	0,6	0,8	0,6	0,6	0,4	0,2	0,6	0,8	0,8	0,6	0	0,4	0,8	1	0,4	0,8	0,6	0,4	0,6	0,8	0,6	0,2	0,4
ZPM	0,4	0,6	0,6	0,8	0,6	0,6	0,6	0,2	0,6	0,4	0,8	1,0	0,4	0,4	0,4	0,4	1	0,6	0,8	0,4	0,2	0,4	0,6	0,6	0,6	0,6
śred.	0,31	0,46	0,53	0,69	0,63	0,56	0,54	0,41	0,50	0,50	0,76	0,90	0,61	0,23	0,34	0,63	0,79	0,61	0,60	0,55	0,30	0,43	0,55	0,56	0,44	0,45
v	46,6	26,0	30,7	21,2	28,3	26,7	32,5	33,0	29,2	20,7	19,7	11,5	27,9	63,9	41,7	25,8	27,0	45,4	27,2	21,0	54,4	23,5	33,9	32,4	41,6	30,4

Rys. 72. Tempo rotacji w imiennictwie na obszarze współczesnych województw (odstępy 5-letnie) (dane z tab. 29)

imiona żeńskie



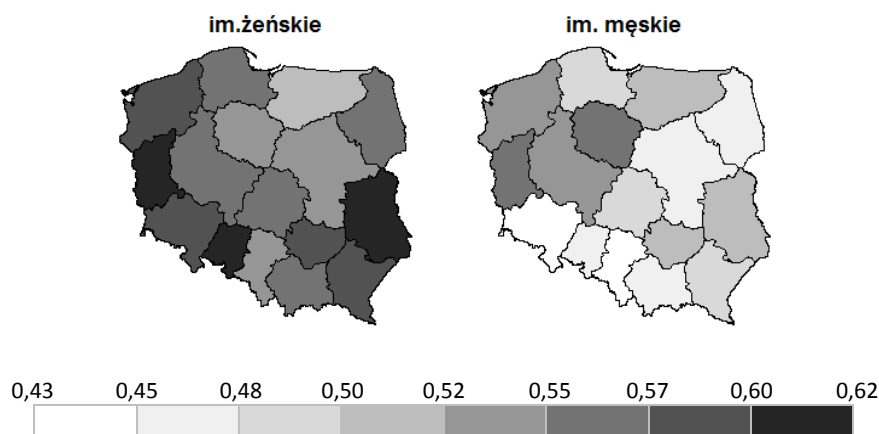
imiona męskie



Tab. 30. Wynik testu istotności różnic między tempem rotacji w imiennictwie żeńskim i męskim na obszarze współczesnych województw (przedziały 5-letnie, 1950–2010)

woj.	średnia wskaźnika tempa rotacji		różnica	wartość p
	im. żeńskie	im. męskie		
DLN	0,58	0,45	0,14	0,069
KUJ	0,54	0,55	-0,02	0,861
LBL	0,60	0,51	0,09	0,190
LBS	0,60	0,55	0,05	0,489
ŁDZ	0,57	0,49	0,08	0,209
MŁP	0,57	0,46	0,11	0,292
MAZ	0,54	0,48	0,06	0,219
OPL	0,62	0,48	0,14	0,121
PDK	0,58	0,49	0,09	0,363
PDL	0,55	0,48	0,08	0,356
POM	0,57	0,49	0,08	0,445
ŚLS	0,54	0,43	0,11	0,131
ŚWK	0,58	0,51	0,08	0,337
WLP	0,55	0,54	0,015	0,870
WRM	0,52	0,51	0,015	0,776
ZPM	0,58	0,54	0,046	0,427

Rys. 73. Średnie wartości wskaźnika tempa rotacji w imiennictwie żeńskim i męskim z okresu 1950–2010 na obszarze współczesnych województw (dane z tab. 30)



Pewne podobieństwa do układu geograficznego często obserwowanego w przypadku parametrów rozkładu częstości (zob. rozdział III) ujawniają się jednak, gdy przyjrzymy się temu, w których województwach odnotowano tendencję rosnącą w zakresie tempa rotacji. Takie zachowanie obserwujemy właściwie tylko w obrębie imiennictwa żeńskiego (w imiennictwie męskim trudno mówić o jakiejś wyraźnej tendencji) na obszarze województw: podkarpackiego i lubelskiego, a także opolskiego, podlaskiego i małopolskiego, a więc – wyjąwszy woj. opolskie – we wschodniej części Polski. Na przeciwnym biegunie

plasują się województwa: dolnośląskie, zachodniopomorskie, mazowieckie, warmińsko-mazurskie i pomorskie, gdzie tempo rotacji nie wykazywało tendencji rosnącej i utrzymywało się na względnie stałym poziomie lub podlegało wahaniom niezależnie od upływającego czasu. Oznacza to tyle, że pierwsza grupa województw wschodnich, początkowo charakteryzująca się wolniejszym tempem rotacji, w następnych latach zrównała się w tym zakresie z pozostałymi województwami. Potwierdza to analiza współczynnika zmienności (v) z tab. 29, zwłaszcza w przypadku imiennictwa żeńskiego – w ostatnich latach rozpatrywanego okresu jest on znacznie niższy niż początkowo, a zatem różnice w rotacji między poszczególnymi województwami uległy zmniejszeniu.

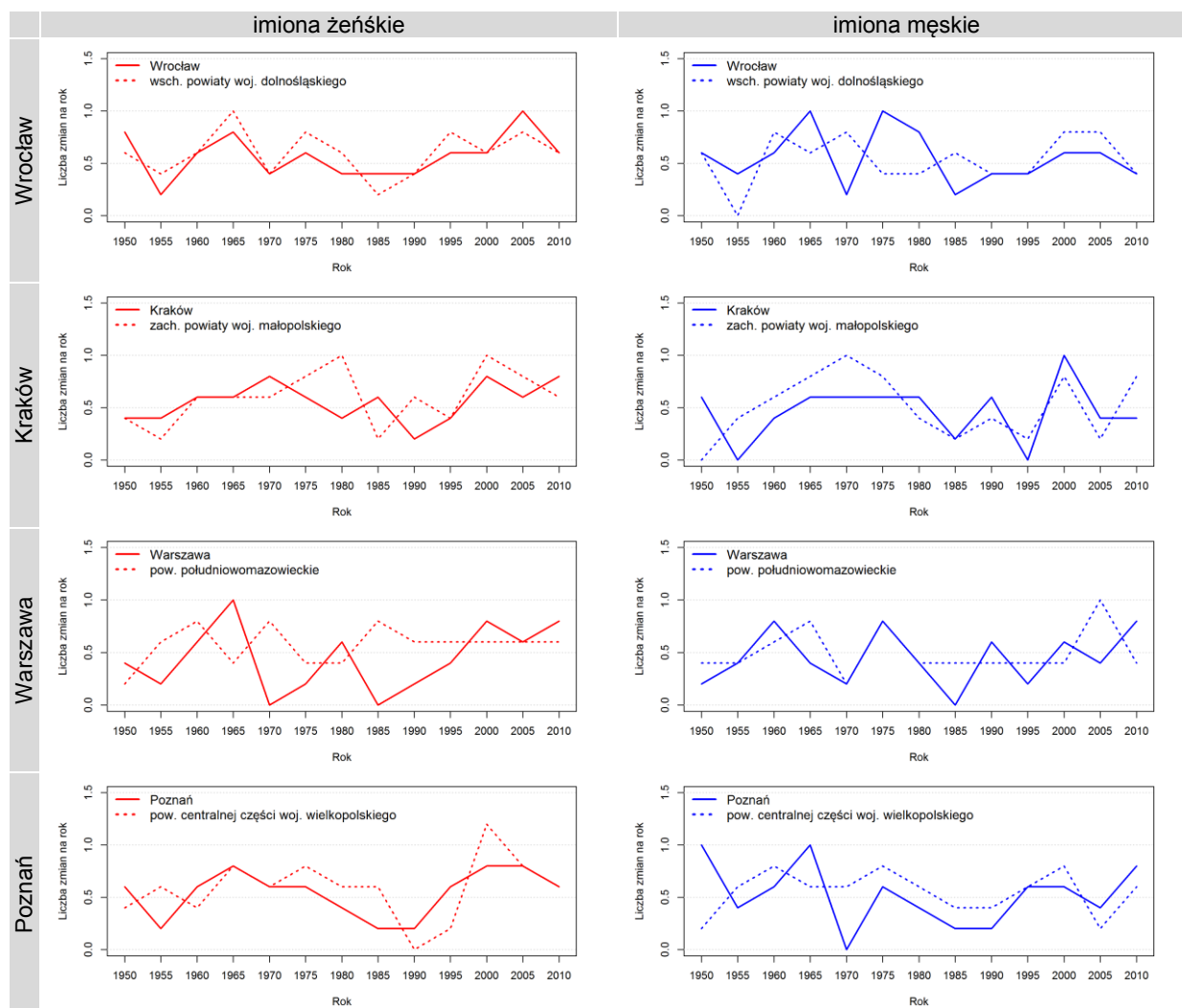
IV.3. Tempo rotacji w imiennictwie obszarów o różnym stopniu urbanizacji

Celem tego badania jest sprawdzenie, czy w zakresie tempa rotacji wśród najczęstszych imion nadawanych przez mieszkańców ośrodków wielkomiejskich – Wrocławia, Krakowa, Warszawy i Poznania – i obszarów o mniejszym stopniu zurbanizowania występowały bądź występują różnice oraz czy ulegały one zmianom na przestrzeni XX i XXI w. Tak jak w poprzednim podrozdziale, badanie ograniczono do zmierzenia tempa rotacji wśród dziesięciu najczęstszych imion w odstępach 5-letnich w okresie 1945–2010. W odróżnieniu od rozdziału poświęconego badaniu parametrów rozkładu częstości, w niniejszej części z miastami porównuję nie cały pozostały obszar województwa, a wybrane powiaty ziemskie w ich pobliżu. Liczba powiatów do porównania była dobierana tak, by liczba osób, którym nadano imię w danym roku była podobna do tej, która występowała w mieście, a w każdym razie duża na tyle, by wielkość słownika (a więc i długość listy frekwencyjnej) w każdym roku wynosiła minimum 100 pozycji. Badaniu poddano następujące pary miast i otaczających je obszarów:

- miasto Wrocław oraz powiaty wschodniej części województwa dolnośląskiego (dzierzoniowski, milicki, oleśnicki, oławski, strzeliński, średzki, trzebnicki, wołowski, wrocławski);
- miasto Kraków oraz powiaty zachodniej części województwa małopolskiego (bocheński, chrzanowski, krakowski, myślenicki, oświęcimski, proszowicki, wadowicki, wielicki)
- miasto Warszawa oraz powiaty południowomazowieckie (garwoliński, grójecki, legionowski, miński, nowodworski, otwocki, piaseczyński, pruszkowski, sochaczewski, warszawski zachodni, wołomiński, żyrardowski);
- miasto Poznań oraz powiaty środkowowielkopolskie (gnieźnieński, grodziski, kościański, nowotomyski, obornicki, poznański, szamotulski, średzki, śremski, wągrowiecki, wrzesiński).

Wyniki obliczeń tempa rotacji dla miast i obszarów poza nimi (osobno dla imion męskich i żeńskich) przedstawiono w formie wykresów na rys. 74. W tab. 31 zestawiono dane o średnim tempie rotacji na poszczególnych obszarach. Dla każdej pary miasto-prowincja w trzecim i czwartym wierszu zamieszczono informacje o różnicy między średnimi i wartości p z testu t -Studenta na ich istotność. Analogicznie w trzeciej i czwartej kolumnie zestawiono dane o różnicy między średnim tempem rotacji między imiennictwem żeńskim i męskim na każdym z obszarów oraz wartości p dla tych różnic.

Rys. 74. Porównanie tempa rotacji w obrębie 10 najczęstszych imion w wybranych ośrodkach wielkomijskich na tle sąsiadujących z nimi obszarów o niższym stopniu urbanizacji (odstępny 5-letnie, 1950–2010)



Na podstawie zamieszczonych powyżej wykresów (rys. 74) możemy wywnioskować, że pomiędzy obszarami miejskimi i otaczającymi je obszarami o niższym stopniu urbanizacji nie występują wyraźne różnice w zakresie rotacji wśród najczęstszych dziesięciu imion z list frekwencyjnych, a z pewnością nie można tu mówić o trwałych i znaczących różnicach, trudno zaobserwować też jakieś wyraźne tendencje. Choć mogłoby się wydawać, że zmiany w imiennictwie powinny zachodzić prędzej w miastach, to dane tego nie potwierdzają, a wręcz mogą dowodzić czegoś odwrotnego. Jeśli przyjrzymy się wynikom średnich wskaźników tempa rotacji dla całego rozpatrywanego okresu, to przekonamy się, że były one niemal zawsze nieznacznie wyższe na obszarze powiatów niż w miastach (z wyjątkiem porównania imiennictwa męskiego we Wrocławiu i na sąsiadujących terenach). Różnice te są jednak mało znaczące – w żadnym przypadku nie możemy mówić o ich istotności (por. wysokie wartości p dla średnich w miastach i powiatach). Zarówno w miastach, jak i

w sąsiadujących z nimi powiatach średnie tempo rotacji było wyższe w imiennictwie żeńskim niż męskim (najwyższa różnica wystąpiła w Krakowie i powiatach małopolskich, brak różnic odnotowano z kolei w Warszawie), jednak w żadnym przypadku również nie były to różnice istotne statystycznie.

Tab. 31. Wyniki testu istotności różnic między wskaźnikami tempa rotacji w miastach i na prowincji oraz między imiennictwem żeńskim i męskim na ich obszarze

	średnie tempo rotacji			
	im. żeńskie	im. męskie	różnica	wartość <i>p</i>
Wrocław	0,57	0,55	0,02	0,829
powiaty	0,60	0,54	0,06	0,472
różnica	-0,03	0,02		
wartość <i>p</i>	0,502	0,877		
Kraków	0,55	0,46	0,09	0,255
powiaty	0,60	0,51	0,09	0,307
różnica	-0,05	-0,05		
wartość <i>p</i>	0,553	0,598		
Warszawa	0,45	0,45	0,00	1,000
powiaty	0,57	0,51	0,06	0,502
różnica	-0,12	-0,06		
wartość <i>p</i>	0,303	0,455		
Poznań	0,54	0,52	0,02	0,837
powiaty	0,58	0,55	0,03	0,721
różnica	-0,05	-0,03		
wartość <i>p</i>	0,534	0,760		

IV.4. Szczegółowe badanie tempa rotacji w skali rocznej – casus II wojny światowej⁶⁹

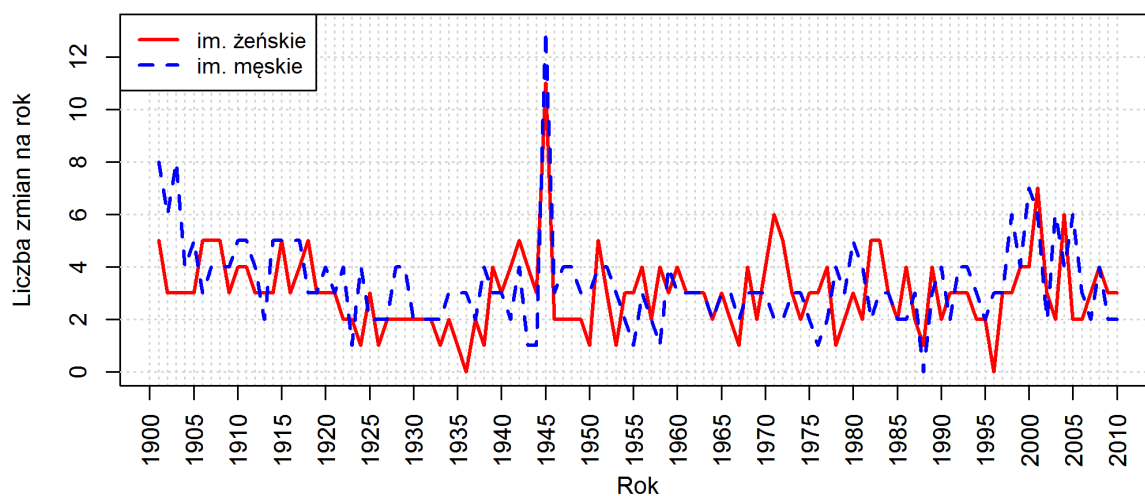
W poprzednich częściach rozdziału tempo rotacji w imiennictwie ogólnopolskim i w poszczególnych regionach mierzone było w pięcio- lub dziesięcioletnich interwałach. Jak już wspomniano, taki sposób pomiaru umożliwia łatwe zaobserwowanie ogólnych tendencji dotyczących tempa wymiany w obrębie najczęstszych imion. Warto jednak przeprowadzić bardziej szczegółowe badanie zjawiska rotacji, które może ujawnić pewne wyraźne zdarzenia, jak się w tej kwestii mogły dokonać. Można bowiem założyć, że niektóre ważne momenty historyczne (jak np. II wojna światowa, zmiany ustroju politycznego, „karnawał” Solidarności itd.) na przestrzeni badanych 110 lat, mogły skutkować gwałtownymi, punktowymi zmianami, które niekoniecznie musiały przynieść efekt długofalowy, jak odnotowana wcześniej tendencja wzrastająca w imiennictwie ogólnopolskim. Już wcześniej w podrozdziale IV.1 zwróciłem uwagę, że badanie zmian wśród 100 najczęstszych imion w stosunku do 5 lat wstecz dla każdego roku (rys. 70) ujawniło znaczny, acz chwilowy wzrost tempa rotacji w latach 40. W niniejszej części prześledzone zostaną zmiany w obrębie najczęstszych imion w następujących po sobie latach.

W badaniu porównywano listy frekwencyjne imion męskich i żeńskich z następujących po sobie lat w obrębie najniższych rang, zaczynając od dziesięciu najczęstszych imion i stopniowo zwiększając tę liczbę o 10. Wyróżniające się na tle całego badanego okresu wartości wskaźnika tempa rotacji zaczęły się zarysowywać dopiero od czwartej dziesiątki najczęstszych imion wzwyż. Przy stu najczęstszych imionach otrzymałem obraz widoczny na wykresie na rys. 75.

Na jego podstawie możemy zaobserwować kilka wyraźnie zarysowujących się „zdarzeń” w zakresie rotacji. Przede wszystkim jest to zdecydowanie niespotykana w innych latach różnica aż 13 imion męskich i 11 żeńskich (spośród 100) między latami 1944 i 1945. Ponadto zwraca uwagę także wzmożona wymiana imion na początku XX w. (męskich), a także przełom XX i XXI w. oraz odnotowany już wcześniej wzrost rotacji wśród imion żeńskich na początku lat 70. Poza okresami wzrostu wymiany imiennictwa mamy też do czynienia z jej znacznymi spadkami, łącznie z brakiem jakichkolwiek zmian (im. żeńskie - 1936 i 1996 r., im. męskie – 1988 r.).

⁶⁹ Fragmenty niniejszego podrozdziału zostały opublikowane w rozszerzonej formie jako artykuł w t. LXI czasopisma „Onomastica” (Swoboda 2017).

Rys. 75. Rotacja w obrębie 100 najczęstszych imion w następujących po sobie latach w skali ogólnopolskiej (lata 1900-2010)



W przypadku nagłego wzrostu tempa rotacji w latach 2000 i 2001 można pokusić się o roboczą interpretację tego faktu. Co prawda, jak mogliśmy zauważyć na wcześniejszych wykresach, generalnie w latach dwutysięcznych tempo rotacji było na wysokim poziomie, mając jednak do czynienia z punktowymi skokami właśnie w tych latach, nie można wykluczyć, że taki „odświeżający” powiew innowacji wśród najczęstszych imion miał związek z wejściem w nowe tysiąclecie. Wymiana imion mogła towarzyszyć licznym symbolicznym działaniom „milenijnym”, które ludzie wówczas podejmowali. Z dużo większą ostrożnością można próbować tak rozumieć wysoką liczbę zmian w obrębie imion męskich na początku XX w. Mogło tu bowiem dojść do zniekształcenia rzeczywistego obrazu, gdyż posiadane przez nas dane dla kilku pierwszych lat są z oczywistych względów znacznie uboższe (zwłaszcza w przypadku imion męskich) niż dla pozostałych lat.⁷⁰ Ponadto tym wysokim wskazaniom rotacji wśród imion męskich nie towarzyszyło (co oczywiście nie jest warunkiem koniecznym) analogicznie zjawisko wśród żeńskich antroponimów.

Nie powinniśmy mieć natomiast wątpliwości, co do zbieżności największego wzrostu w zakresie wymiany imion z rokiem 1945. Jest to data zakończenia II wojny światowej i wydaje się, że właśnie z tym wydarzeniem powinniśmy wiązać tak wyraziste zjawisko nazewnictwa, jakim jest wymiana kilkunastu najczęściej nadawanych imion w stosunku do roku poprzedniego. Wymaga ono jednak szerszego zbadania i omówienia, gdyż sam fakt historyczny nie może (i nie musi) być wystarczającym wyjaśnieniem zaobserwowanego

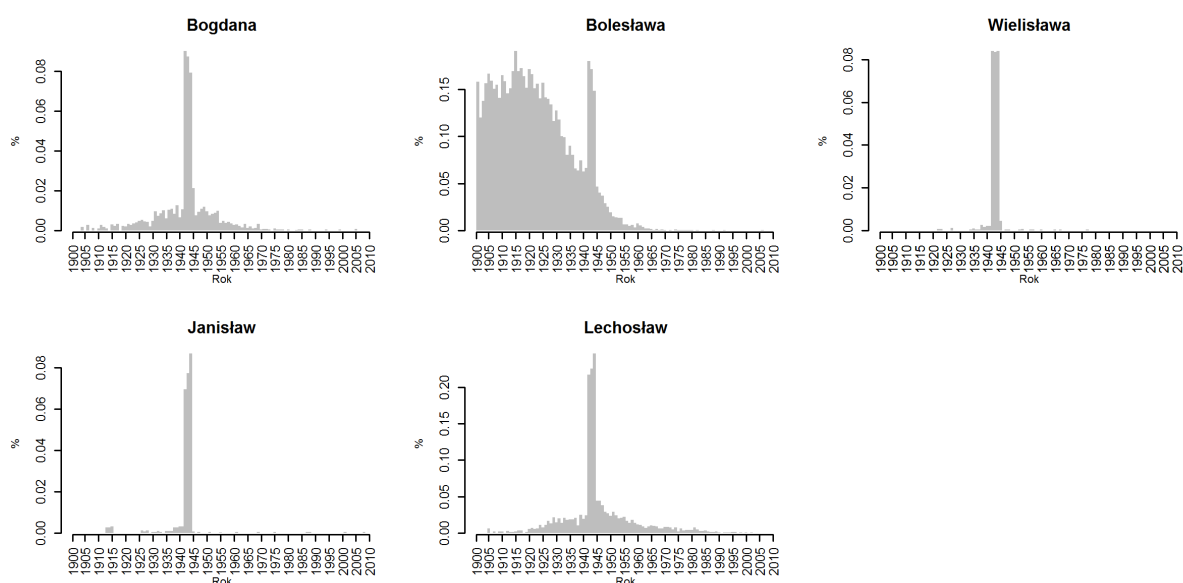
⁷⁰ Jak mogliśmy zaobserwować na rys. 1 w rozdziale I.3.1 poświęconym bazie materiałowej niniejszej rozprawy, liczba osób w najwcześniejszych latach znacznie odbiega od późniejszych okresów.

zjawiska. Jak wiadomo ze studium „Język polski czasu drugiej wojny światowej” pod redakcją I. Bajerowej (1996), okres wojny odcisnął silne piętno na ówczesnej polszczyźnie, włączając w to także sferę onimiczną. E. Rudnicka-Fira (1996) opisała w nim wpływ języka niemieckiego oraz niemieckiej polityki językowej na przemiany w nazewnictwie, w tym w antroponimii. Na podstawie danych z ksiąg urodzeń dla wybranych miejscowości przedstawiła, w jaki sposób nie tylko germanizowano imiona Polaków, ale także wpływno, przy pomocy narzędzi administracyjnych, na nadawanie dzieciom narodowości polskiej imion niemieckich bądź genetycznie niemieckich (Pomorze, Śląsk), lub odwrotnie – imion „typowo” polskich w celu napiętnowania i dyskryminacji osób narodowości polskiej (region łódzki). O imionach nadawanych łodzianom urodzonym podczas okupacji hitlerowskiej szczegółowo pisała Elżbieta Umińska-Tytoń (1988) w kontekście wydanego przez administrację tzw. Kraju Warty (niem. Warthegau, Wartheland) rozporządzenia, zalecającego nadawanie polskim dzieciom imion słowiańskich ze specjalnie sporządzonego wykazu.

Imiona męskie, które w 1945 w skali ogólnopolskiej zostały wyparte to: *Brunon, Ernest, Erwin, Eryk, Ginter, Helmut, Herbert, Janisław, Lechosław, Manfred, Walter, Werner, Wilhelm*; żeńskie zaś to: *Adelajda, Bogdana, Bolesława, Edeltrauda, Edyta, Eryka, Gizela, Helga, Hildegarda, Karolina, Wielisława*. Większość z nich to imiona pochodzenia germańskiego. Spadek ich popularności po II wojnie światowej można tłumaczyć ówczesnym negatywnym stosunkiem do wszelkich przejawów niemieckości. Wśród wymienionych imion możemy dostrzec jednak także antroponimy słowiańskie: *Janisław, Lechosław, Bogdana, Bolesława, Wielisława*. Na liście stu najczęstszych imion zostały one zastąpione przez następujące imiona: męskie – *Albin, Arkadiusz, Augustyn, Benedykt, Daniel, Florian, Hieronim, Jakub, Konstanty, Leopold, Marcin, Teofil, Walerian*; żeńskie: *Anastazja, Bernadeta, Felicja, Honorata, Iwona, Longina, Ryszarda, Teodora, Teodozja, Waleria, Zenona*. Jak łatwo zauważyć nie występują wśród nich imiona słowiańskie. Oczywiście spadek popularności imion rodzimych nie musiał mieć związku z wydarzeniami II wojny światowej i mógł być efektem naturalnej wymiany zasobu najczęstszych imion, jednak gdy spojrzymy na wykresy frekwencji względnej wybranych imion, które zostały wyparte z pierwszej setki (rys. 76), to związek ten wydaje się ewidentny.

Tylko w przypadku imienia żeńskiego *Bolesława* możemy mówić o stopniowym spadku popularności w pierwszej połowie XX w. Wszystkie imiona łączy natomiast gwałtowny wzrost nadań w czasie II wojny światowej. Jak wiadomo z pracy E. Umińskiej-Tytoń (1988), hitlerowskie władze okupacyjne wprowadziły rozporządzenie nakazujące nadawanie osobom

Rys. 76. Frekwencja względna imion słowiańskich, które wypadły z listy 100 najczęstszych imion w 1945 r. (lata 1901–2010)



narodowości polskiej imion polskich, które będą łatwo ich identyfikować, dlatego też w łódzkich księgach urodzeń z 1943 r. badaczka odnotowała liczne nadania imion *Kazimierz*, *Kazimiera*, czy zapomniane imiona *Dobromira*, *Mściśław*, natomiast brak w nich było wielu popularnych przed i po wojnie imion jak *Andrzej*, *Barbara*, *Jan*, *Krystyna*, *Maria* itp. Przepis ten jednak został wprowadzony tylko w Kraju Warty,⁷¹ natomiast rozporządzenie nie było nigdy wprowadzone na pozostałych obszarach włączonych do III Rzeszy, nie wydano też podobnych przepisów dla Generalnego Gubernatorstwa (Umińska-Tytoń 1988: 121–123). Warto zatem sprawdzić, czy wyjątkowe zjawisko wzmożonej rotacji w roku 1945 miało miejsce na obszarze całego kraju, czy też dotyczyło to jedynie tych jego części, które w czasach okupacji zostały włączone do Kraju Warty.

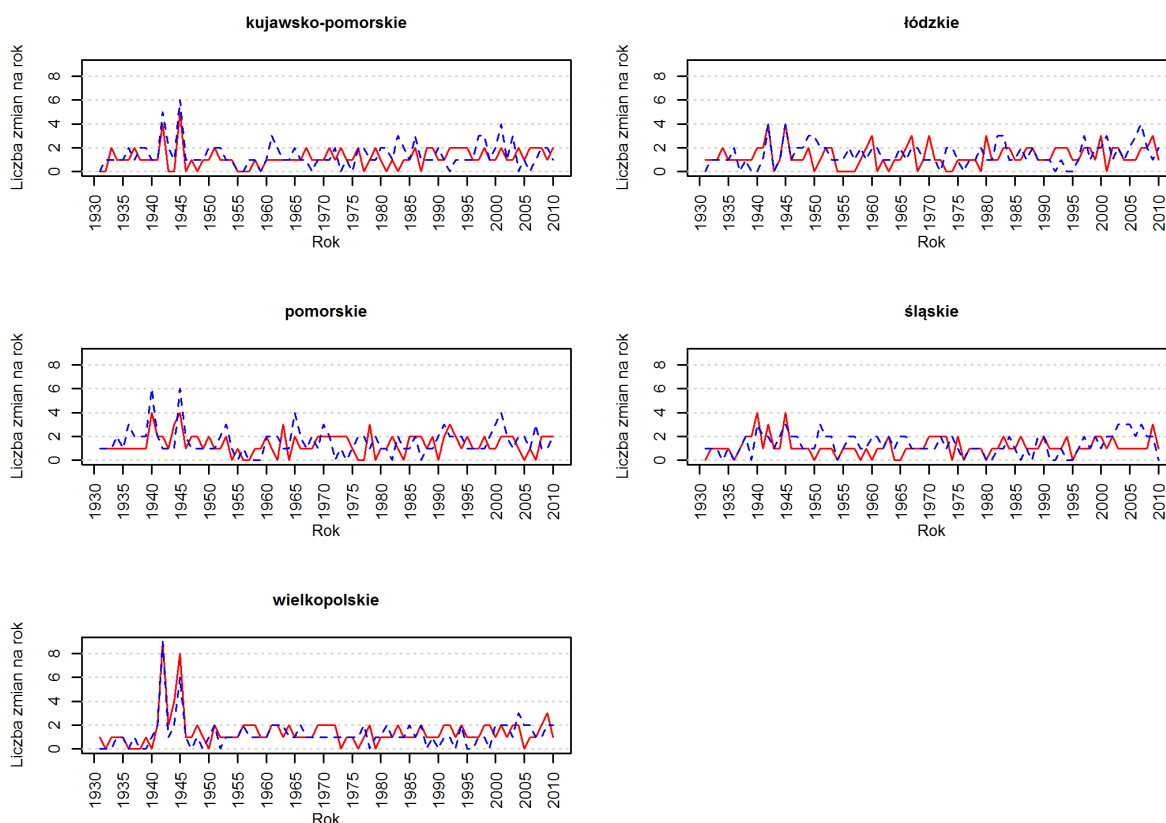
Po przeprowadzeniu badania rotacji w skali rocznej w okresie 1930–2010 na obszarze poszczególnych współczesnych województw⁷² (posłużyłem się tym razem liczbą jedynie 20 najczęstszych imion z uwagi na mniejszą liczbę okazów w skali wojewódzkiej niż ogólnokrajowej), okazało się, że wpływ na zaobserwowaną powyżej wzmożoną rotację

⁷¹ Kraj Warty (niem. Warthegau, Wartheland, Reichsgau Wartheland) – nazwa regionu administracyjnego utworzonego podczas II wojny światowej na terytorium Rzeczypospolitej Polskiej anektowanym przez III Rzeszę, w skład którego weszło międzywojenne województwo poznańskie, większa (zachodnia) część woj. łódzkiego oraz południowo-zachodnia część woj. warszawskiego (do linii Wisły). Obszar ten pokrywa się z większością współczesnego woj. wielkopolskiego, południową częścią woj. kujawsko-pomorskiego oraz zachodnią częścią woj. łódzkiego.

⁷² Z wyjątkiem obszarów, dla których nie dysponowałem wystarczającą ilością danych, włączonych do Polski w całości dopiero po 1945 r. (woj. dolnośląskie, lubuskie, opolskie, warmińsko-mazurskie, zachodniopomorskie).

wywarły zmiany w najczęstszym imiennictwie na obszarze współczesnych województw: kujawsko-pomorskiego, łódzkiego, pomorskiego, śląskiego i wielkopolskiego, co przedstawiono na poniższych wykresach (rys. 77), które w innych województwach się nie ujawniły.

Rys. 77. Liczba zmian obrębie 20 najczęstszych imion w następujących po sobie latach na obszarze współczesnych województw: kujawsko-pomorskiego, łódzkiego, pomorskiego, śląskiego i wielkopolskiego (lata 1930-2010)



Wynika z nich jednak także, że gwałtowny wzrost liczby wymian z roku 1945 był tam poprzedzony niemal równy liczbowo zmianami w roku 1940 – w woj. pomorskim i śląskim oraz 1942 – w woj. kujawsko-pomorskim, łódzkim oraz wielkopolskim, których współczesny obszar, jak już wspomniano, pokrywa się w dużej mierze z obszarem Kraju Warty. Natomiast wszystkie wymienione obszary łączy to, iż poza wschodnią częścią woj. łódzkiego i północno-wschodnią częścią województwa śląskiego, nie wchodziły w obręb Generalnego Gubernatorstwa. By nie czynić zbyt daleko idących uogólnień, warto prześledzić jak kształtowały się zmiany nie w kontekście współczesnego podziału administracyjnego, ale na obszarze poszczególnych stref okupacyjnych okresu II wojny światowej. Na potrzeby tego badania, na podstawie kodów TERC przypisałem w bazie danych każdej osobie urodzonej w okresie 1939–1945 przynależność terytorialną do określonej niemieckiej strefy

okupacyjnej⁷³. W tabelach 32 i 33 zamieszczono informacje o liczbie zmian wśród 20 najczęstszych imion męskich i żeńskich nadawanych w poszczególnych strefach okupacyjnych w kolejnych latach II wojny światowej.

Tab. 32. Liczba zmian na liście 20 najczęstszych imion żeńskich w stosunku do poprzedzającego roku na obszarze poszczególnych stref okupacyjnych

strefa okupacyjna	1939	1940	1941	1942	1943	1944	1945
Gdańsk-Prusy Zachodnie	0	3	3	1	0	1	3
Generalne Gubernatorstwo	0	0	1	0	1	1	1
Górny Śląsk	2	4	0	1	1	1	2
Prusy Wschodnie	1	1	1	1	0	1	2
Kraj Warty	1	0	2	8	1	1	7
Okręg Białostocki	1	1	2	1	1	1	1

Tab. 33. Liczba zmian na liście 20 najczęstszych imion męskich w stosunku do poprzedzającego roku na obszarze poszczególnych stref okupacyjnych

strefa okupacyjna	1939	1940	1941	1942	1943	1944	1945
Gdańsk-Prusy Zachodnie	0	4	2	1	1	1	6
Generalne Gubernatorstwo	0	1	2	1	1	0	0
Górny Śląsk	1	3	2	2	1	1	4
Prusy Wschodnie	0	1	1	1	1	1	1
Kraj Warty	0	0	1	9	1	1	7
Okręg Białostocki	1	2	1	1	1	2	3

Jak łatwo zauważyć, największą skalę osiągnęły zmiany w imiennictwie w Kraju Warty. Pierwsza fala zmian miała tam miejsce w 1942 r., a więc w rok po wprowadzeniu przepisów nakazujących nadawanie Polakom imion słowiańskich z określonego zbioru. Na liście 20 najczęstszych imion doszło do wymiany niemal połowy pozycji – 8 imion żeńskich i 9 męskich. Z nieco mniejszą liczbą zmian, jednak również zauważalną, mamy do czynienia w przypadku prowincji Gdańsk-Prusy Zachodnie i Górny Śląsk. Tutaj wzmożenie rotacji w obrębie najczęstszych imion miało miejsce nieco wcześniej, bo w latach 1940–41. W Prusach Wschodnich, a także we włączonym w 1941 r. do nich Okręgu Białostockim

⁷³ Badaniu poddano jedynie te części poszczególnych obszarów administracyjnych III Rzeszy lub jej stref okupacyjnych, które przed II wojną światową znajdowały się i obecnie znajdują się na terytorium Polski. Poza wcześniej omówionym Krajem Warty były to następujące obszary: Gdańsk-Prusy Zachodnie (Reichsgau Danzig-Westpreussen), do których włączono obszar przedwojennego woj. pomorskiego; Górny Śląsk (Provinz Oberschlesien), do której włączono przedwojenne autonomiczne woj. śląskie oraz zachodnie powiaty woj. krakowskiego i fragment południowo-zachodniej części woj. kieleckiego; Prusy Wschodnie (Provinz Ostpreussen), do których włączono północną część woj. warszawskiego, rejon Działdowa oraz Suwałk i Sejn; Okręg Białystok, do 1941 r. roku będący częścią Białoruskiej SSR, włączony następnie w obręb Prus Wschodnich (współczesne woj. podlaskie stanowi jego zachodnią część); Generalne Gubernatorstwo, na którego obszarze znajdowały się pozostałe części współczesnej Polskie.

zmiany te nie miały tak wyraźnego i gwałtownego przebiegu, a w przypadku Generalnego Gubernatorstwa są ledwie zauważalne. Zwróćmy także uwagę, że w miejscach, gdzie odnotowano znaczne wzmożenie rotacji w latach 1940–42, w 1945 r. doszło do kolejnej fali wymiany imiennictwa. Jeśli chodzi o prowincję Górny Śląsk, należy jednak pamiętać, że do jej obszaru włączono nie tylko przedwojenne autonomiczne województwo śląskie, ale także część zachodnich powiatów przedwojennego województwa krakowskiego (dawna Galicja) oraz fragment południowo-wschodniej części przedwojennego województwa kieleckiego (dawne Królestwo Kongresowe). Po uwzględnieniu tych uwarunkowań historycznych uzyskamy nieco bardziej szczegółowy obraz rotacji w obrębie najczęstszego imiennictwa na obszarze prowincji Górny Śląsk. Informacje o wskaźniku rotacji w każdej z części składowych tego obszaru zamieszczono w tabelach 34 i 45.

Tab. 34. Liczba zmian na liście 20 najczęstszych imion żeńskich w stosunku do poprzedzającego roku na obszarze prowincji Górny Śląsk z uwzględnieniem przedwojennego podziału administracyjnego

	1939	1940	1941	1942	1943	1944	1945
Dawne woj. krakowskie	2	1	1	0	0	1	2
Dawne woj. kieleckie	0	0	0	0	1	1	1
Dawne autonomiczne woj. śląskie	1	6	0	2	1	0	6

Tab.35. Liczba zmian na liście 20 najczęstszych imion męskich w stosunku do poprzedzającego roku na obszarze prowincji Górny Śląsk z uwzględnieniem przedwojennego podziału administracyjnego

	1939	1940	1941	1942	1943	1944	1945
Dawne woj. krakowskie	1	1	1	1	2	1	1
Dawne woj. kieleckie	1	1	1	1	1	1	2
Dawne autonomiczne woj. śląskie	1	6	3	2	3	2	6

Z danych zamieszczonych w tab. 34 i 35 dobitnie wynika, że do wyraźnych i gwałtownych zmian w zakresie rotacji imiennictwa doszło jedynie w śląskiej części prowincji Górny Śląsk, natomiast na obszarze dawnych województw kieleckiego i krakowskiego, podobnie jak w sąsiadującym z nim Generalnym Gubernatorstwie, nie odnotowano znacznej wymiany.

Przyjrzyjmy się teraz, jakich imion dotyczyły zmiany na obszarach poszczególnych niemieckich stref okupacyjnych, w których odnotowano znaczny wzrost tempa rotacji. W tabelach 36–41 zamieszczono listy 20 najczęstszych imion w kolejnych latach z przedziału 1939–45. Pogrubioną i pochyloną czcionką oznaczono te imiona, które w następującym roku wypadły z listy najczęstszych, natomiast kolorem szarym wyróżniono komórki tabeli, w których znalazły się imiona, które w danym roku awansowały do pierwszej dwudziestki.

Tab. 36. Listy frekwencyjne 20 najczęstszych imion żeńskich w prowincji Gdańsk-Prusy Zachodnie

	1939	1940	1941	1942	1943	1944	1945
1	Krystyna	Krystyna	Krystyna	Krystyna	Krystyna	Krystyna	Krystyna
2	Maria	Jadwiga	Jadwiga	Jadwiga	Irena	Urszula	Maria
3	Teresa	Maria	Maria	Maria	Maria	Jadwiga	Teresa
4	Jadwiga	Irena	Irena	Irena	Jadwiga	Irena	Barbara
5	Irena	Teresa	Gertruda	Urszula	Urszula	Maria	Irena
6	Barbara	Helena	Urszula	Gertruda	Teresa	Teresa	Halina
7	Zofia	Zofia	Teresa	Helena	Gertruda	Barbara	Jadwiga
8	Helena	Gertruda	Helena	Teresa	Zofia	Renata	Urszula
9	Halina	Elżbieta	Elżbieta	Elżbieta	Elżbieta	Helena	Zofia
10	Janina	Anna	Zofia	Zofia	Helena	Elżbieta	Janina
11	Anna	Barbara	Anna	Anna	Anna	Gertruda	Danuta
12	Urszula	Urszula	Barbara	Barbara	Barbara	Anna	Elżbieta
13	Elżbieta	Halina	Halina	Renata	Renata	Zofia	Helena
14	Danuta	Janina	Małgorzata	Halina	Wanda	Ewa	Anna
15	Stefania	Wanda	Eryka	Wanda	Halina	Halina	Ewa
16	Stanisława	Stefania	Wanda	Eryka	Ewa	Wanda	Wanda
17	Genowefa	Małgorzata	Janina	Janina	Eryka	Regina	Regina
18	Marianna	Danuta	Renata	Małgorzata	Janina	Lidia	Marianna
19	Gertruda	Eryka	Hildegarda	Ewa	Małgorzata	Janina	Genowefa
20	Wanda	Ewa	Regina	Regina	Regina	Eryka	Renata

Tab. 37. Listy frekwencyjne 20 najczęstszych imion męskich w prowincji Gdańsk-Prusy Zachodnie

	1939	1940	1941	1942	1943	1944	1945
1	Jan	Jan	Jan	Henryk	Henryk	Henryk	Jan
2	Stanisław	Henryk	Henryk	Jan	Jan	Ryszard	Jerzy
3	Jerzy	Ryszard	Ryszard	Ryszard	Ryszard	Jan	Kazimierz
4	Henryk	Jerzy	Jerzy	Jerzy	Jerzy	Jerzy	Henryk
5	Józef	Józef	Józef	Józef	Józef	Józef	Stanisław
6	Kazimierz	Stanisław	Gerard	Edmund	Edmund	Edmund	Józef
7	Edward	Kazimierz	Roman	Franciszek	Zygmunt	Edward	Ryszard
8	Ryszard	Roman	Edmund	Roman	Edward	Zygmunt	Zbigniew
9	Tadeusz	Zygmunt	Edward	Gerard	Franciszek	Roman	Tadeusz
10	Franciszek	Franciszek	Stanisław	Zygmunt	Roman	Franciszek	Marian
11	Zygmunt	Edmund	Kazimierz	Edward	Gerard	Gerard	Andrzej
12	Andrzej	Gerard	Franciszek	Kazimierz	Stanisław	Stanisław	Roman
13	Roman	Edward	Zygmunt	Stanisław	Kazimierz	Andrzej	Zygmunt
14	Marian	Tadeusz	Paweł	Zygfryd	Stefan	Zygfryd	Edmund
15	Edmund	Andrzej	Brunon	Brunon	Brunon	Stefan	Mieczysław
16	Zbigniew	Paweł	Zygfryd	Paweł	Waldemar	Kazimierz	Stefan
17	Stefan	Brunon	Stefan	Waldemar	Andrzej	Waldemar	Franciszek
18	Mieczysław	Stefan	Andrzej	Stefan	Paweł	Brunon	Zdzisław
19	Czesław	Leon	Waldemar	Leon	Zygfryd	Bernard	Edward
20	Eugeniusz	Marian	Tadeusz	Andrzej	Bernard	Leon	Czesław

Tab. 38. Listy frekwencyjne 20 najczęstszych imion żeńskich na obszarze dawnego woj. śląskiego w prowincji Górny Śląsk

	1939	1940	1941	1942	1943	1944	1945
1	Maria	Maria	Maria	Maria	Maria	Maria	Maria
2	Krystyna	Krystyna	Krystyna	Krystyna	Krystyna	Krystyna	Krystyna
3	Helena	Anna	Anna	Urszula	Anna	Anna	Irena
4	Irena	Elżbieta	Irena	Anna	Urszula	Urszula	Urszula
5	Anna	Helena	Helena	Irena	Irena	Irena	Helena
6	Elżbieta	Irena	Elżbieta	Elżbieta	Helena	Helena	Anna
7	Teresa	Małgorzata	Urszula	Helena	Elżbieta	Elżbieta	Teresa
8	Jadwiga	Urszula	Małgorzata	Renata	Małgorzata	Renata	Janina
9	Małgorzata	Jadwiga	Renata	Małgorzata	Renata	Teresa	Jadwiga
10	Aniela	Hildegarda	Jadwiga	Jadwiga	Teresa	Małgorzata	Stefania
11	Janina	Teresa	Teresa	Hildegarda	Jadwiga	Jadwiga	Halina
12	Łucja	Róża	Hildegarda	Gertruda	Gertruda	Barbara	Barbara
13	Stefania	Renata	Gertruda	Teresa	Barbara	Stefania	Elżbieta
14	Urszula	Gertruda	Róża	Stefania	Stefania	Joanna	Małgorzata
15	Halina	Łucja	Łucja	Róża	Hildegarda	Hildegarda	Wanda
16	Gertruda	Eryka	Stefania	Barbara	Łucja	Róża	Danuta
17	Danuta	Adelajda	Lidia	Łucja	Róża	Gertruda	Łucja
18	Zofia	Stefania	Adelajda	Adelajda	Joanna	Lidia	Zofia
19	Bronisława	Lidia	Eryka	Joanna	Adelajda	Łucja	Joanna
20	Lidia	Barbara	Barbara	Elfryda	Lidia	Adelajda	Aniela

Tab. 39. Listy frekwencyjne 20 najczęstszych imion męskich na obszarze dawnego woj. śląskiego w prowincji Górny Śląsk

	1939	1940	1941	1942	1943	1944	1945
1	Józef	Józef	Henryk	Henryk	Henryk	Henryk	Józef
2	Jan	Jan	Jan	Józef	Józef	Józef	Jan
3	Henryk	Henryk	Józef	Jan	Jan	Jan	Henryk
4	Jerzy	Jerzy	Jerzy	Jerzy	Jerzy	Jerzy	Jerzy
5	Stanisław	Rudolf	Paweł	Franciszek	Piotr	Piotr	Stanisław
6	Edward	Franciszek	Franciszek	Paweł	Franciszek	Paweł	Marian
7	Franciszek	Paweł	Rudolf	Piotr	Rudolf	Franciszek	Andrzej
8	Eugeniusz	Edward	Gerard	Rudolf	Paweł	Rudolf	Antoni
9	Paweł	Antoni	Edward	Edward	Gerard	Antoni	Tadeusz
10	Zygmunt	Stanisław	Ryszard	Gerard	Antoni	Edward	Franciszek
11	Antoni	Gerard	Alfred	Ryszard	Alfred	Gerard	Piotr
12	Alojzy	Ryszard	Antoni	Alfred	Karol	Karol	Alojzy
13	Kazimierz	Eryk	Alojzy	Karol	Edward	Andrzej	Zygmunt
14	Czesław	Alfred	Piotr	Antoni	Ryszard	Alfred	Paweł
15	Marian	Karol	Karol	Alojzy	Alojzy	Ryszard	Ryszard
16	Andrzej	Alojzy	Ginter	Ginter	Andrzej	Stefan	Kazimierz
17	Roman	Roman	Eugeniusz	Eryk	Stefan	Alojzy	Stefan
18	Tadeusz	Andrzej	Eryk	Stefan	Eryk	Krzysztof	Edward
19	Stefan	Stefan	Stanisław	Herbert	Stanisław	Norbert	Roman
20	Adolf	Adolf	Andrzej	Eugeniusz	Zygmunt	Stanisław	Eugeniusz

Tab. 40. Listy frekwencyjne 20 najczęstszych imion żeńskich w Kraju Warty

	1939	1940	1941	1942	1943	1944	1945
1	Maria	Maria	Maria	Janina	Janina	Janina	Maria
2	Krystyna	Krystyna	Krystyna	Kazimiera	Jadwiga	Jadwiga	Krystyna
3	Teresa	Zofia	Janina	Jadwiga	Kazimiera	Bożena	Barbara
4	Zofia	Teresa	Zofia	Bożena	Bożena	Kazimiera	Irena
5	Marianna	Irena	Teresa	Stanisława	Stanisława	Stanisława	Janina
6	Janina	Janina	Irena	Czesława	Czesława	Mirostawa	Halina
7	Barbara	Marianna	Jadwiga	Mirosława	Mirosława	Czesława	Teresa
8	Irena	Barbara	Barbara	Bronisława	Zdzisława	Zdzisława	Zofia
9	Jadwiga	Jadwiga	Marianna	Zdzisława	Bronisława	Bronisława	Jadwiga
10	Anna	Anna	Halina	Bogumiła	Bogumiła	Krystyna	Anna
11	Helena	Helena	HeLena	Krystyna	Krystyna	Maria	Danuta
12	Halina	Halina	Anna	Maria	Maria	Bogumiła	Marianna
13	Stanisława	Stanisława	Stanisława	Irena	Wiesława	Wiesława	Wanda
14	Danuta	Danuta	Kazimiera	Wiesława	Irena	Irena	Helena
15	Genowefa	Wanda	Danuta	Zofia	Teresa	Zofia	Urszula
16	Kazimiera	Kazimiera	Wanda	Anna	Anna	Teresa	Elżbieta
17	Wanda	Elżbieta	Urszula	Halina	Halina	Anna	Stanisława
18	Urszula	Genowefa	Elżbieta	Bolesława	Zofia	Barbara	Genowefa
19	Stefania	Urszula	Władystawa	Teresa	Bolesława	Halina	Ewa
20	Elżbieta	Stefania	Czesława	PeLagia	Danuta	Danuta	Bożena

Tab. 41. Listy frekwencyjne 20 najczęstszych imion męskich w Kraju Warty

	1939	1940	1941	1942	1943	1944	1945
1	Jan	Jan	Jan	Kazimierz	Stanisław	Stanisław	Jan
2	Stanisław	Stanisław	Stanisław	Stanisław	Kazimierz	Jerzy	Stanisław
3	Józef	Józef	Józef	Jerzy	Jerzy	Kazimierz	Jerzy
4	Marian	Kazimierz	Kazimierz	Czesław	Czesław	Czesław	Józef
5	Kazimierz	Henryk	Jerzy	Zdzisław	Zdzisław	Zdzisław	Henryk
6	Henryk	Marian	Henryk	Bogdan	Bogdan	Bogdan	Andrzej
7	Tadeusz	Jerzy	Marian	Zbigniew	Zbigniew	Zbigniew	Marian
8	Jerzy	Tadeusz	Tadeusz	Bronisław	Bronisław	Bronisław	Tadeusz
9	Andrzej	Ryszard	Władysław	Włodzimierz	Włodzimierz	Włodzimierz	Ryszard
10	Ryszard	Andrzej	Ryszard	Jan	Mirosław	Mirosław	Kazimierz
11	Władysław	Władysław	Czesław	Mirosław	Jan	Wojciech	Zdzisław
12	Edward	Stefan	Stefan	Wojciech	Wojciech	Jan	Zbigniew
13	Czesław	Czesław	Andrzej	Bolesław	Bolesław	Bolesław	Stefan
14	Stefan	Zygmunt	Zdzisław	Władysław	Józef	Henryk	Mieczysław
15	Eugeniusz	Edward	Zygmunt	Bogusław	Lechosław	Józef	Zygmunt
16	Zygmunt	Eugeniusz	Edward	Józef	Władysław	Lechosław	Władysław
17	Mieczysław	Franciszek	Zbigniew	Lechosław	Mieczysław	Ryszard	Czesław
18	Franciszek	Zenon	Mieczysław	Henryk	Henryk	Mieczysław	Zenon
19	Zdzisław	Zdzisław	Eugeniusz	Ryszard	Ryszard	Tadeusz	Eugeniusz
20	Zenon	Mieczysław	Zenon	Bogumił	Bogusław	Władysław	Edward

Zgodnie z przytaczanymi wcześniej obserwacjami E. Rudnickiej-Firy i E. Umińskiej-Tytoń, widzimy, że w przypadku Kraju Warty po 1941 r. odnotowano znaczny wzrost nadań imion niemal wyłącznie słowiańskich, które wyparły imiona o innej proveniencji, także germańskiej. Odwrotnie było na pozostałych obszarach włączonych do Rzeszy: Górnym Śląsku i Pomorzu – tam nowe antroponimy na liście dwudziestu najczęstszych imion stanowią w głównej mierze imiona związane z kulturą niemiecką.⁷⁴

O tym, że nadawanie większości z nich było związane z niemiecką polityką językową, świadczą dwa fakty. Pierwszym z nich jest szybkość, w jakim awansowały do grupy najczęstszych 20 imion – nie było to stopniowe, „naturalne” zwiększanie popularności, ale nagły skok. Drugim zaś, jeszcze bardziej przekonującym argumentem jest to, że w 1945 r. gwałtownie straciły na popularności i zostały na powrót zastąpione przez imiona, które były przez nie wyparte w pierwszych latach wojny. Widzimy zatem, że antroponimy nadawane częściej w początkowych latach okupacji, niezależnie od ich genetycznego pochodzenia (słowiańskiego czy germańskiego) bądź ich związków z kulturą polską lub niemiecką, w takim samym stopniu były wypierane w roku 1945 r., zapewne dlatego, że były kojarzone z opresyjną polityką okupacyjną.

⁷⁴ Należy podejrzewać, że wiele imion wprowadzonych na Pomorzu i Górnym Śląsku miało formę i pisownię niemiecką (Rosa, Erika, Erich, Richard, Karl...), natomiast już po zakończeniu wojny dostosowano je do pisowni polskiej. Jak podają A. Dziubok i B. Linek, na Górnym Śląsku do końca 1948 r. wpłynęło ok. 280 tys. wniosków o zmianę lub „sprostowanie” imienia lub nazwiska (Dziubok, Linek 2011: 275). Zmianę umożliwił Dekret z dnia 10 listopada 1945 r. o zmianie i ustalaniu imion i nazwisk (Dz. U. nr 56, poz. 1310). Dokument ten umożliwił m.in. zmianę imienia w wypadku, gdy miało ono „brzmienie niepolskie” (art. 10, ust. 2, pkt 1).

V. Badanie regionalnego zróżnicowania frekwencji względnej imion metodą hierarchicznej analizy skupień

Analiza wskaźników liczbowych dotyczących rozkładu częstości imiennictwa w Polsce, przedstawiona w rozdziale III, pozwoliła na zaobserwowanie pewnych prawidłowości w ich rozmieszczeniu geograficznym. W niniejszej części przyjrę się temu, czy występują bądź występowały charakterystyczne konfiguracje geograficzne związane z układem imion na listach frekwencyjnych na poszczególnych obszarach oraz czy konfiguracje te ulegały na przestrzeni lat przeobrażeniom. Kształt tych list warunkowany jest przez frekwencję względną poszczególnych imion. Wiadomo, że na niektórych obszarach stosunkowo często nadawane są imiona, które w imiennictwie ogólnopolskim są spotykane rzadziej (zob. Swoboda 2013: 56–67). Nawiązując do terminu zaproponowanego przez K. Handke (1986/1987), możemy określić je mianem antroponimicznych regionalizmów frekwencyjnych. Ale nawet najbardziej popularne imiona w skali ogólnopolskiej mogą występować na różnych obszarach z różnym natężeniem. Przykładowo w kilku powojennych latach najczęstszym imieniem męskim w Polsce był *Jan*, jednak na obszarze czterech współczesnych województw (lubelskiego, małopolskiego, podkarpackiego i świętokrzyskiego) częściej występowało wówczas imię *Stanisław*. Takich różnic w układzie list frekwencyjnych imion na różnych terenach może być o wiele więcej (wspominałem o nich już w rozdziale II.1.3).

Wykonanie ręcznych analiz porównawczych wielu list frekwencyjnych z poszczególnych obszarów byłoby bardzo trudne i czasochłonne. By tego uniknąć, w celu zbadania zróżnicowania regionalnego w zakresie frekwencji względnej imion nadawanych na obszarze 16 współczesnych województw w następujących po sobie latach z przedziału 1946–2010 (rozdział V.5) oraz na obszarze współczesnych powiatów województwa śląskiego w przedziałach 20-letnich w okresie 1931–2010 (rozdział V.6) posłużyłem się metodą hierarchicznej analizy skupień. Metoda ta jest szeroko stosowana m. in. w badaniach stylometrycznych (zob. np. Rybicki 2011, Rybicki, Heydel 2013, Eder 2014, 2017) do graficznego zobrazowania podobieństw między wieloma tekstami różnych autorów na podstawie porównania frekwencji użycia przez nich określonej liczby np. najczęstszych wyrazów, kolokacji czy własności składniowych. Narzędzie to zostało także wykorzystane do zbadania zróżnicowania regionalnego imiennictwa w Stanach Zjednoczonych w przywoływanej już pracy włoskich badaczy (Barucca i in. 2015), która w dużej mierze stanowiła inspirację dla niniejszego rozdziału.

V.1. Wprowadzenie

Jak pisze A. Balicki, „Analizę skupień (ang. *cluster analysis*) możemy określić jako dział wielowymiarowej analizy statystycznej obejmujący zbiór technik znajdujących zastosowanie w badaniu danych o obiektach wielowymiarowych (jednostek lub zmiennych) w celu podzielenia zbioru tych obiektów na grupy (skupienia), które w świetle danych wykazują odrębność. (...) Ideą analizy skupień jest (...) znajdowanie jednorodnych podzbiorów w niejednorodnym zbiorze obiektów. Mówiąc inaczej, chodzi o wyłonienie grup obiektów, które są bardziej podobne do obiektów tworzących daną grupę niż do obiektów spoza tej grupy” (Balicki 2009: 208–209). Głównym celem analizy skupień jest więc ułożenie obiektów (reprezentowanych przez określone cechy) w grupy w taki sposób, aby odzwierciedlić stopień ich powiązania. Badanie to ma charakter stricte eksploracyjny, co oznacza, że nie są określane żadne zewnętrzne kryteria kategoryzacji obiektów, ani też nie zakłada się z góry liczby skupień, które będą uzyskane, choć oczywiście, bazując na wcześniejszych badaniach parametrów rozkładu częstości możemy podejrzewać, jaki może być ich układ.

W analizie skupień każdy obiekt charakteryzowany jest przez daną liczbę cech X_1, X_2, \dots, X_p , których wartości możemy przedstawić w formie wektorów wierszowych $x' = (x_{i1}, x_{i2}, \dots, x_{ip})$ $i = 1, 2, \dots, n$ stanowiących wiersze w macierzy danych \mathbf{X} (Balicki: 212–213). W moich badaniach, mówiąc o obiektach, mam na myśli zbiory imion nadawanych na pewnych obszarach geograficznych (np. w województwach, powiatach itd.) w danym przedziale czasowym, zaś cechy tych obiektów to same imiona i ich frekwencja względna. Macierz danych, o której mowa, jest więc po prostu zbiorem list frekwencyjnych imion nadawanych na danym obszarze, który może prezentować się następująco (tab. 42):

Tab. 42. Fragment przykładowej macierzy danych zawierającej wartości cech (frekwencję względną imion) reprezentujących obiekty (województwa)

	MARIA	KRYSTYNA	ANNA	IRENA	HELENA	JANINA	...
dolnośl.	4.34	6.34	4.00	3.31	3.11	5.50	...
kuj.-pom.	11.22	4.91	6.48	3.77	6.74	4.57	...
lubelskie	10.94	4.89	10.91	3.53	9.53	3.05	...
lubuskie	3.26	6.04	4.49	3.48	3.51	5.76	...
łódzkie	8.78	5.28	4.15	3.29	3.62	1.59	...
małopol.	3.35	6.20	3.15	4.57	3.92	6.36	...
...

Cechy opisujące obiekty, nazywane też cechami diagnostycznymi, wybierane są zwykle z większego zbioru cech. Do kwestii tego, ile i jakie cechy (imiona) zostaną wybrane, wróć

w podrozdziale V.3. Ich układ na listach frekwencyjnych (podyktowany procentowym udziałem w zbiorze imion nadawanych na danym obszarze) będzie decydował o stopniu zróżnicowania pomiędzy poszczególnymi obiektami.

Istnieją dwie podstawowe techniki grupowania (klastrowania) obiektów wielowymiarowych, mianowicie podział i hierarchia (podtypy: aglomeracyjny i deglomeracyjny). Najczęściej stosowane i w większości wypadków wystarczające jest hierarchiczne grupowanie aglomeracyjne (Balicki 2009: 211–212), toteż zdecydowałem się właśnie na jego zastosowanie, co zostało zasygnalizowane w tytule rozdziału. Istota tej techniki zasadza się na tym, „że wyjściowo każdy element zbioru Ω jest traktowany jako odrębna grupa jednoelementowa. Opierając się na miarach odległości lub podobieństwa przeprowadza się sekwencyjne grupowanie (czyli aglomerację) obiektów, w zależności od tego, jak dalece różnią się one między sobą lub na ile są podobne. Pierwsze skupienie zawsze się tworzy z połączenia dwóch obiektów najbardziej podobnych. W kolejnych etapach bądź to dołącza się pojedynczy obiekt do istniejącego już skupienia bądź też łączy skupienia, również według najmniejszej odległości lub największego podobieństwa. (...) W efekcie wielokrotnego powtórzenia postępowania – łączenia dwóch najbliższych skupień, przeliczanie odległości itd. – otrzymujemy jedną ostateczną grupę obiektów” (Balicki: 211–212).

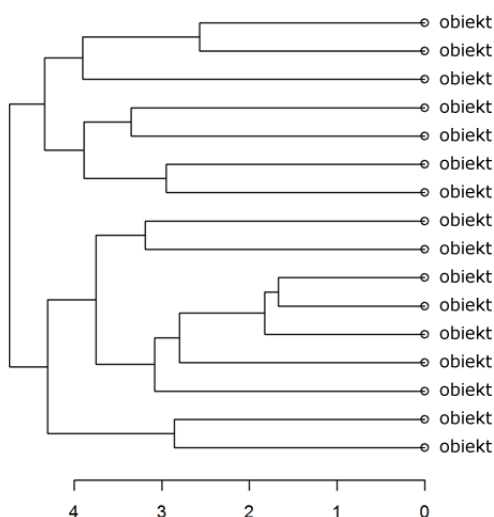
W przypadku naszych badań procedury aglomeracyjne hierarchicznej analizy skupień wymagają następujących kroków. W pierwszej kolejności stworzona zostanie lista frekwencyjna r najczęstszych imion w całym zbiorze w danym roku (osobno dla imion męskich i żeńskich) oraz listy frekwencyjne dla każdego z porównywanych obiektów (województwa). Następnie zostanie sporządzona macierz odległości⁷⁵ dla każdej pary obiektów w oparciu o frekwencję względną (procentową) każdej ze zmiennych (imion z listy r najczęstszych imion). Macierz odległości (niepodobieństw) przypomina znane z atlasów drogowych tablice odległości między poszczególnymi miastami, w których dla każdej pary miejscowości podawana jest odległość np. w kilometrach. Istnieje cały szereg miar, które stosuje się przy budowie macierzy odległości – kwestia wyboru konkretnej metryki zostanie szerzej omówiona w podrozdziale V.2).

⁷⁵ Niezależnie od wybranej techniki grupowania w analizie skupień kluczową rolę odgrywają pojęcia odległości i podobieństwa, których miary obliczane są w oparciu o zaobserwowane wartości cech badanych obiektów. Wynik klasteryzacji zależy od tego, w jaki sposób postanowimy mierzyć odległość (lub podobieństwo) między parami lub grupami obiektów. Biorąc pod uwagę, że wartości cech, którymi dysponujemy, to dane ilościowe (frekwencje względne imion), będziemy stosować miary odległości, wykorzystywane przy pomiarze zróżnicowania obiektów według cech ilościowych (Balicki 2003: 214).

Poza wyborem miary odległości, dzięki której stworzymy macierz odległości, stanowiącą podstawę analizy skupień, musimy także dokonać wyboru metody wiązania (łączenia, grupowania), która pozwoli na określenie odległości między kolejnymi skupieniami, powstałymi z powiązanych obiektów. Metod łączenia, podobnie jak miar odległości, jest wiele – kryteria wyboru jednej z nich przedstawię w podrozdziale V.4.

Efektem końcowym hierarchicznej analizy skupień jest jej obraz graficzny w postaci dendrogramu, czyli wykresu w formie drzewa binarnego (rys. 78). Zbudowane jest ono z „gałęzi” (kładów, węzłów), odpowiadających skupieniom, oraz z „liści”, które reprezentują poszczególne obiekty (np. teksty czy zbiory imion). Dendrogram prezentuje relacje podobieństwa (lub niepodobieństwa) pomiędzy tymi obiektami. Gałęzie mogą mieć jeden liść lub więcej liści, które umieszczone są na poziomie zerowym, natomiast węzły dendrogramu znajdują się na wysokości, która odpowiada mierze odległości (niepodobieństwa) między poszczególnymi skupieniami. Długość gałęzi na której znajdują się dwa obiekty (lub większe skupienia) odzwierciedla stopień, w jakim są one ze sobą zbieżne. Im dłuższa gałąź, tym różnice pomiędzy dwoma obiektami są większe. Na wstępie każdy obiekt stanowi odrębne skupienie (wysokość odcięcia na poziomie $h = 0$), następnie na podstawie obliczonych odległości (miar niepodobieństwa) między listami frekwencyjnymi najczęstszych imion tworzy kolejne, większe skupienia wraz z obiektami, z którymi dzieli najmniejszą odległość.

Rys. 78. Dendrogram – graficzne przedstawienie wyników hierarchicznej analizy skupień



V.2. Wybór metryki odległości

Jak już wspomniałem, istnieje szerokie spektrum miar odległości, które stosuje się w procedurze budowy macierzy odległości między obiektami.⁷⁶ Każda z nich jest stosowana w zależności od potrzeb badawczych. Jak pisze A. Balicki, miary odległości (d) są funkcjami wartości cech obiektu X_1, \dots, X_p

$$d_{rs} = f(x_r, x_s)$$

opartymi na odległościach między punktami r i s w przestrzeni wielowymiarowej. Każdy z obiektów jest punktem w przestrzeni cech, w której odległość dwóch punktów jest wartością metryki p -wymiarowej przestrzeni (X_1, \dots, X_p) (Balicki 2009: 214).

W niniejszej pracy zdecydowałem się na zastosowanie metryki euklidesowej (odległości euklidesowej), uznawanej za klasyczną i najbardziej popularną.⁷⁷ Jest to najczęściej wybierana miara odległości, gdy badacz nie dysponuje założeniami dotyczącymi podobieństwa lub jego braku pomiędzy poszczególnymi obiektami, a tak jest właśnie w przypadku materiału imienniczego. W odróżnieniu od badań stylometrycznych (bazujących na materiale dzieł literackich, w których przy pomocy korpusu testowego możemy dobrać metody pod kątem tego, czy w prawidłowy sposób oddają podobieństwo tekstów tego samego autora), w przypadku naszego materiału nie możemy a priori stwierdzić, które obiekty (zbiory imion z różnych regionów) powinny wykazywać podobieństwo. Stwierdzenie istnienia bądź nieistnienia tego podobieństwa jest dopiero celem naszego badania.

Metryka euklidesowa jest określana jako odległość euklidesowa dwóch punktów w przestrzeni p -wymiarowej, czyli odległość geometryczną między obiektami w przestrzeni, którą, mówiąc obrazowo, mogliśmy zmierzyć linijką. Wyrażana jest następującym wzorem (za: Balicki 2009: 215):

⁷⁶ Są to m.in. metryka miejska (city block, Manhattan metric, taxicab), metryka dominacji (Czebyszewa), metryka minimum, Canberra, metryka Minkowskiego, współczynnik Lance'a i Williamsa, współczynnik Czekanowskiego (miara Matusity), współczynnik dywergencji, metryka Gowera (Balicki 2009: 214–222). W pakiecie statystycznym R, w którym wykonano wszystkie obliczenia, wbudowana funkcja *dist* (Distance Matrix Computation) standardowo obsługuje metryki: euklidesową, maximum (Czebyszewa), Manhattan, Canberra oraz odległość Minkowskiego (R Core Team 2015). W opracowanym na potrzeby badań stylometrycznych pakiecie *stylo* poza podstawowymi metrykami dostępne są również te, stworzone specjalnie na potrzeby badań tekstów literackich, takie jak: Classic Delta (delta Burrowsa), Argamon's Linear Delta, Eder's Delta, Eder's Simple (Eder i in. 2013).

⁷⁷ Użyli jej także Barucca i in. (2015) w swojej pracy na temat dystrybucji geograficznej imion w Stanach Zjednoczonych.

$$d_{rs} = \sqrt{\sum_{j=1}^p (x_{rj} - x_{sj})^2}$$

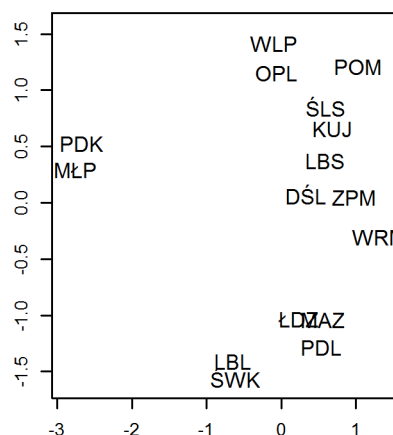
gdzie j - poszczególne zmienne (cechy, czyli imiona), p - liczba zmiennych. Jest to pierwiastek z sumy kwadratów różnic wartości wszystkich zmiennych (cech diagnostycznych) dla obiektów r i s . Dla każdej pary obiektów wyliczane są zatem różnice pomiędzy wartościami każdej z cech (czyli frekwencjami względnymi wybranych imion). Macierz odległości to tabelaryczne zestawienie odległości między każdą parą obiektów. Fragment przykładowej macierzy odległości między obiektami i nadawanymi w nich imionami zamieszczono w poniższej tabeli 43.

Tab. 43. Fragment przykładowej macierzy odległości między województwami w zakresie układu imion na listach frekwencyjnych

	dolnośl.	kuj.-pom.	lubelskie	lubuskie	łódzkie	...
kuj.-pom.	3.92					
lubelskie	3.70	3.94				
lubuskie	4.26	3.39	3.36			
łódzkie	3.54	3.43	2.55	2.88		
małop.	3.36	3.38	4.42	3.70	3.93	
...

Odległości między obiektami można zwizualizować, stosując technikę skalowania wielowymiarowego (ang. *multidimensional scaling*). Jest to metoda ekstrakcji cech na podstawie macierzy odległości (lub niepodobieństwa) między obiektami, jej celem jest wyznaczenie współrzędnych w nowym układzie współrzędnych, tak by odległości w tym układzie były podobne do oryginalnych dystansów w macierzy odległości (Biecek,

Rys. 79. Skalowanie wielowymiarowe na podstawie macierzy odległości euklidesowych dla 16 współczesnych województw w oparciu o frekwencję względną 100 najczęstszych imion żeńskich nadanych w 1946r.



Trajkowski 2011: 10). Graficzną ilustracją skalowania wielowymiarowego jest biplot (dwuwymiarowy). Na rys. 79 przedstawiono przykładowy wynik skalowania wielowymiarowego odległości pomiędzy obszarami współczesnych województw (w zakresie 100 najczęściej nadawanych imion żeńskich) w 1946 r.

V.3. Dobór cech diagnostycznych

Cechami obiektów – rozumianych tu jako np. regiony czy też zbiory imion nadawanych na ich obszarze – są same imiona odznaczające się odpowiednią frekwencją. Macierz odległości, która jest podstawą hierarchicznej analizy skupień, może opierać się na wszystkich imionach odnotowanych na poszczególnych obszarach bądź też na ich wyborze. Na pierwszy rzut oka może się wydawać, że tylko wzięcie pod uwagę wszystkich dostępnych danych może dać jedyne wiarygodne wyniki. Mając, jednak na uwadze, że zbiory imion, nazwisk czy teksty języka naturalnego cechują się wysokim stopniem koncentracji i specyficznym rozkładem częstości, gdzie bardzo niewielka część jednostek opisuje znakomitą większość zbioru (por. rys. 5 w rozdziale II.1.3), można założyć, że pewne wyraźne i stabilne konfiguracje przestrzenne, które zamierzam zbadać, możemy uzyskać przy użyciu mniejszej liczby cech (zmiennych).

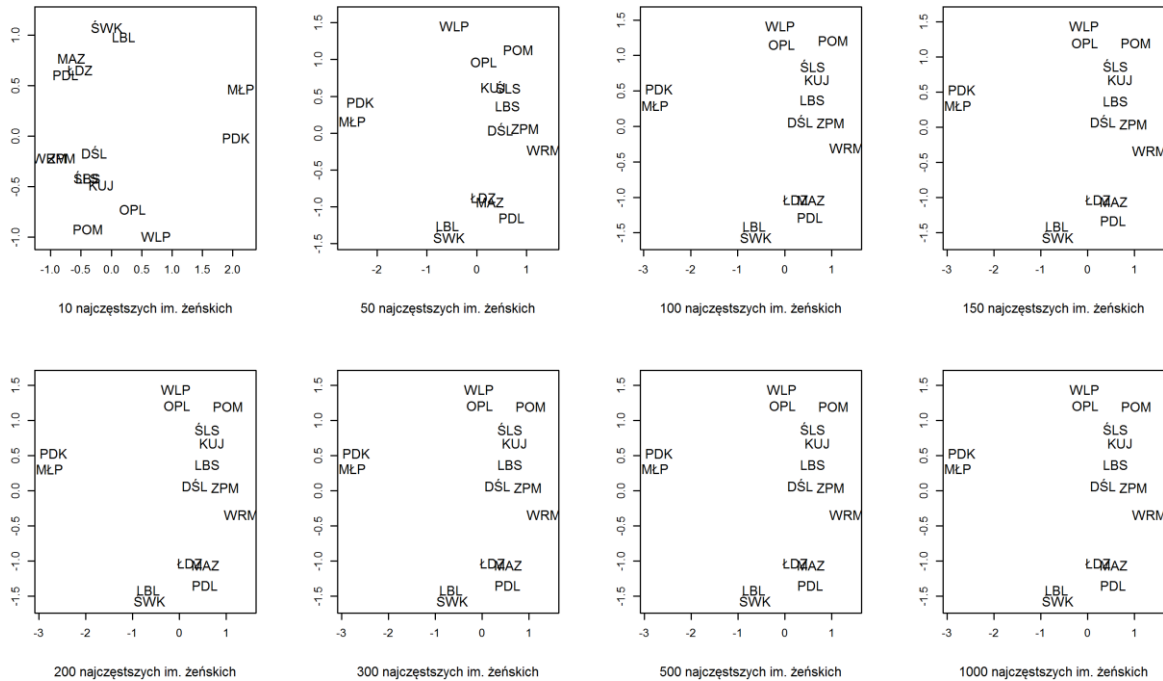
W celu ustalenia, jaka liczba najczęstszych imion jest wystarczająca do uzyskania stabilnych wyników analizy skupień, sporządziłem macierze odległości dla dużego zbioru imion obejmującego okres 1946–2010, dobierając kolejno 10, 50, 100, 150, 200, 300, 500 i 1000 najczęstszych imion, oraz dla zbioru mniejszego z roku (1946 r.), dobierając kolejno 10, 20, 40, 60, 80, 100, 120 i 150 imion. Na podstawie każdej z nich wykonałem skalowanie wielowymiarowe, które pozwoliło mi porównać wyniki dla tak zróżnicowanych list oraz przekonać się, jaka jest minimalna liczba najczęstszych imion, na podstawie której otrzymamy stabilne wyniki. Różnice w liczbie dobieranych najczęstszych imion podyktowane są tym, że o ile dla całego okresu w każdym województwie dysponujemy odpowiednio długimi słownikami imion, to już w konkretnym roku liczba nadanych imion oraz ich słownik będą odpowiednio mniejsze.

Wyniki skalowania wielowymiarowego przedstawione zostały na rysunkach 80–83. Na podstawie porównania poniższych wykresów dla obiektów n -elementowych (tzn. reprezentowanych przez n najczęstszych imion) zarówno w przypadku dużego (rys. 80 i 81), jak i mniejszego zbioru (rys. 82 i 83) można stwierdzić, że na poziomie 100 pierwszych imion z list frekwencyjnych wyniki są stabilne i zwiększenie liczby badanych cech nie wpływa w znaczącym stopniu na zmianę konfiguracji obiektów na wykresach, a więc na odległości między nimi.⁷⁸ Wydaje się więc zasadne, by na potrzeby analizy dotyczącej

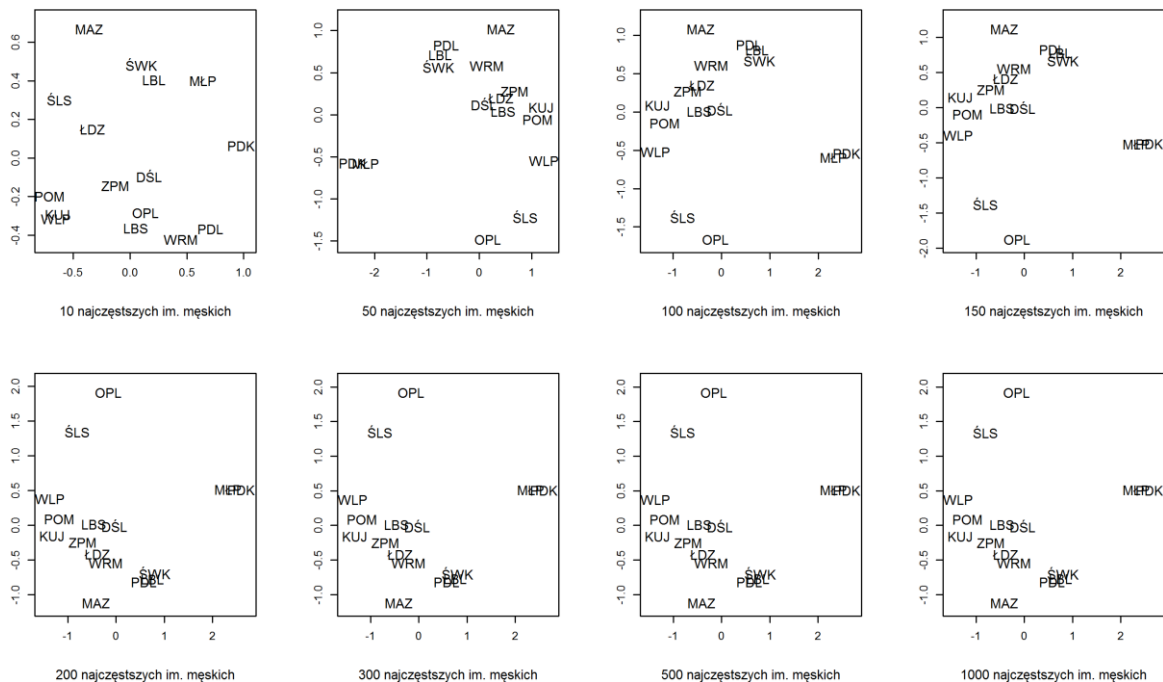
⁷⁸ Choć na rys. 82 dwa ostatnie panele (dla 120 i 150 imion) na pierwszy rzut oka różnią się diametralnie, to w rzeczywistości mamy tu do czynienia z lustrzanym odbiciem – odległości pomiędzy poszczególnymi obiektami nie uległy zmianie.

obszaru województw liczbę cech diagnostycznych (najczęstszych imion) ograniczyć do stu, co znacząco ułatwi jej przeprowadzenie.

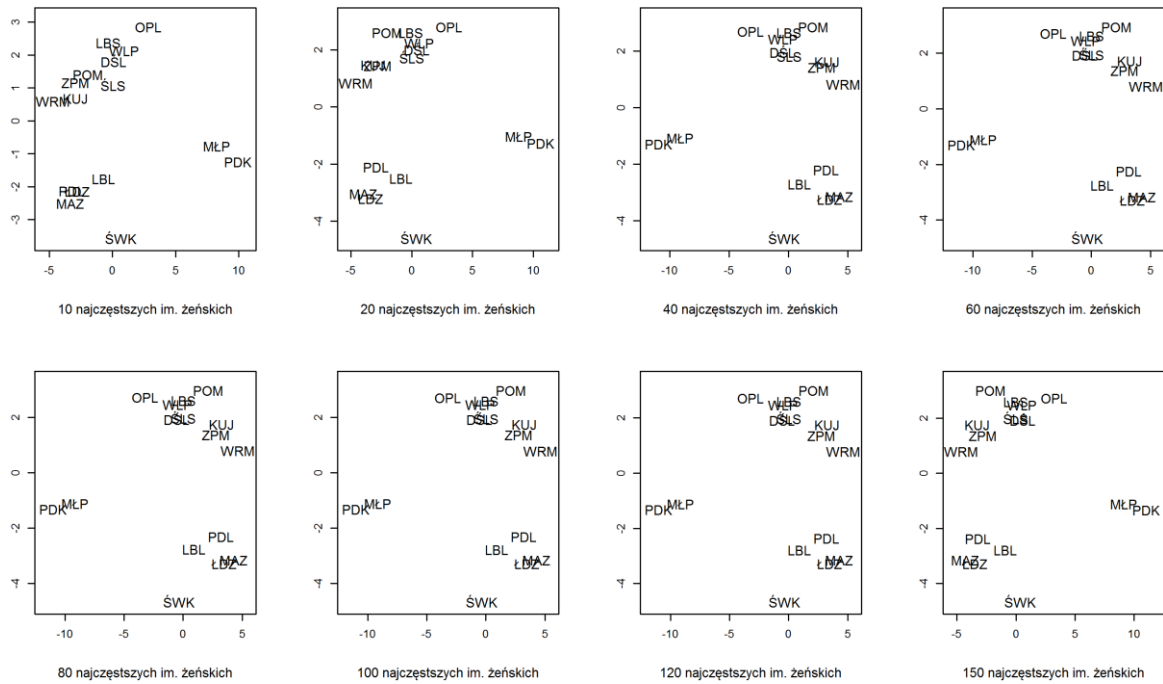
Rys. 80. Skalowanie wielowymiarowe na podstawie macierzy odległości euklidesowych stworzonych w oparciu o frekwencję względną r najczęstszych imion żeńskich w poszczególnych województwach w okresie 1946–2010



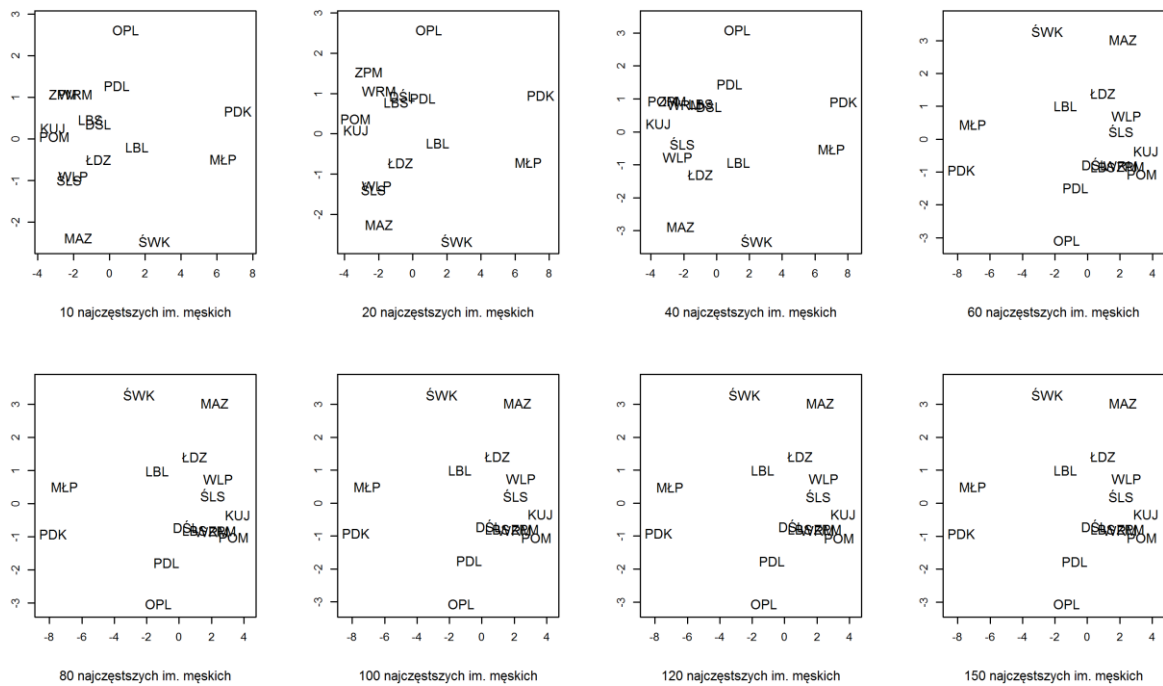
Rys. 81. Skalowanie wielowymiarowe na podstawie macierzy odległości euklidesowych stworzonych w oparciu o frekwencję względną r najczęstszych imion męskich w poszczególnych województwach w okresie 1946–2010



Rys. 82. Skalowanie wielowymiarowe na podstawie macierzy odległości euklidesowych stworzonych w oparciu o frekwencję względną r najczęstszych imion żeńskich w poszczególnych województwach w 1946 r.



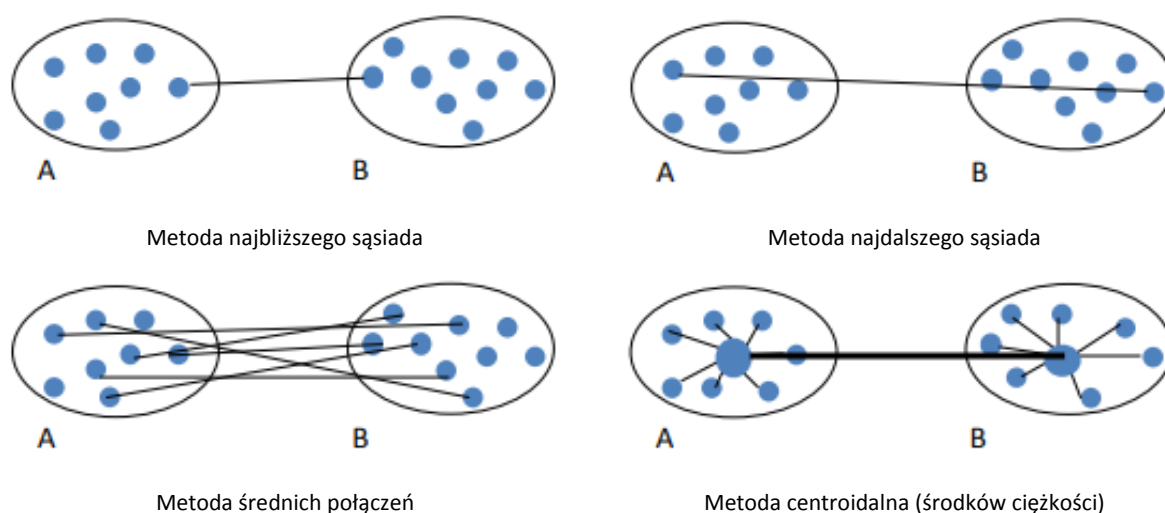
Rys. 83. Skalowanie wielowymiarowe na podstawie macierzy odległości euklidesowych stworzonych w oparciu o frekwencję względną r najczęstszych imion męskich w poszczególnych województwach w 1946 r.



V.4. Wybór metody łączenia

Poza wyborem metryki odległości, dzięki której tworzymy macierz odległości, stanowiącą podstawę analizy skupień, musimy także dokonać wyboru metody łączenia (wiązania, grupowania), pozwalającej na określenie odległości między nowymi skupieniami, powstałymi z powiązanych obiektów. Jak już wspomniano, istnieje wiele różnych metod łączenia. Do najczęściej wykorzystywanych należą: metoda najbliższego sąsiada (single linkage method), metoda najdalszego sąsiada (complete linkage), metoda średniej grupowej/średnich połączeń (UPGMA, unweighted pair-group method using arithmetic averages), metoda centroidalna/środków ciężkości (UPGMC – unweighted pair-group method using the centroid average), metoda mediany/ważonych środków ciężkości (WPGMC – weighted pair-group method using the centroid average) oraz metoda Warda (Balicki 2009: 261–279).

Rys. 84. Graficzna ilustracja wybranych metod wiązania w hierarchicznej analizie skupień (oprac. na podstawie mat. dydaktycznych M. M. Długosz, AGH: http://home.agh.edu.pl/~mmd/_media/dydaktyka/adp/analiza_skupien.pdf)



Wybór metody łączenia nie jest kwestią trywialną, należy bowiem wybrać taką, która możliwie poprawnie odzwierciedli „realne” połączenia między obiektami, które wynikają z macierzy odległości. Jednak wielu autorów, wykorzystujących w swoich badaniach metodę hierarchicznej analizy skupień, wybiera po prostu jedną z dostępnych w używanym przez nich oprogramowaniu statystycznym metod, nie bacząc na to, czy jest ona odpowiednia dla opracowywanych danych (Saraçli i in. 2013: 203).

Zastosowanie konkretnej metody łączenia może opierać się na badaniach empirycznych, jak w przypadku badań stylometrycznych⁷⁹ – jeśli wiemy jak poszczególne obiekty (np. teksty różnych autorów) powinny się ze sobą łączyć, wówczas wybieramy metodę, która najbardziej odpowiada stanowi faktycznemu⁸⁰. Jeśli jednak nie możemy *a vista* ocenić, który typ łączenia jest najbardziej skuteczny (a tak jest w przypadku naszego badania o charakterze eksploracyjnym), istnieją też statystyczne metody mierzenia dopasowania dendrogramu (będącego wynikiem z zastosowania wybranej procedury grupowania hierarchicznego) do macierzy odległości.⁸¹ W niniejszej pracy zdecydowałem się wykorzystać w tym celu metodę współczynnika korelacji kofenetycznej (cophenetic correlation coefficient, dalej CCC) z uwagi na jej prawdopodobnie największe rozpowszechnienie (McGarigal i in. 2000: 117; Tan i in. 2005: 544, Carteron i in. 2012: 64, Saraçlı i in. 2013: 203) oraz na stosunkową łatwość jej zastosowania. W celu oszacowania CCC oblicza się współczynnik korelacji liniowej Pearsona pomiędzy wejściową macierzą odległości **D** (tą, którą otrzymałem w poprzednim etapie, zob. V.2), a macierzą dendrogramu **C** (tj. macierzą kofenetyczną). Elementy macierzy **C** to odległości odpowiadające poziomom łączenia, na których pary obiektów łączą się w tym samym skupieniu po raz pierwszy. Współczynnik korelacji kofenetycznej wyrażany jest następująco:

$$r_{kof.} = \frac{\left[\frac{n(n-1)}{2} \sum_{r<s} d_{rs} \cdot c_{rs} - \sum_{r<s} d_{rs} \cdot \sum_{r<s} c_{rs} \right]}{\sqrt{\left[n(n-1)/2 \sum_{r<s} d_{rs}^2 - \left(\sum_{r<s} d_{rs} \right)^2 \right] \left[n(n-1)/2 \sum_{r<s} c_{rs}^2 - \left(\sum_{r<s} c_{rs} \right)^2 \right]}}$$

gdzie: n - liczba obiektów, $r, s \in \{1, \dots, n\}$, d_{rs} - elementy macierzy odległości między obiektami **D**, c_{rs} - elementy macierzy kofenetycznej **C** (za: Balicki 2009: 284).

W zależności od opracowania za wiarygodne odzwierciedlenie przez zbudowany dendrogram podobieństw lub różnic między obiektami uznaje się wtedy, gdy wartości współczynnika korelacji kofenetycznej dla danej metody wiązania wynoszą powyżej 0,75 (McGarigal i in. 2000: 117) lub nawet 0,8 (Balicki 2009: 284–285) (w przedziale [-1,1]).

⁷⁹ O dyskusji na temat wyboru metody łączenia w badaniach stylometrycznych zob. Eder 2017: 51–52.

⁸⁰ Wiedząc o tym, że teksty tego samego autora są do siebie najbardziej podobne, możemy łatwo ocenić, czy wybrana metoda łączenia dobrze odzwierciedla to podobieństwo na tle tekstów innych autorów.

⁸¹ Są to m.in. współczynnik korelacji kofenetycznej, wskaźnik całościowego dopasowania, statystyka Gowera, metoda STRESS Kruskala, współczynnik Randa (Balicki 2009: 283–291; Machowska-Szewczyk, Sompolska-Rzechuła 2012: 162). Wszystkie oparte są na porównaniu wyjściowej macierzy odległości i macierzy odległości na dendrogramie, które traktuje się jako poziomy, na których jednostki pojawiają się po raz pierwszy w skupieniu (Balicki 2013: 283–284).

Zasadniczo jednak należy przyjąć, że im wyższa wartość CCC (dla danej metody w porównaniu z innymi), tym lepsze odwzorowanie (Machowska-Szewczyk, Sompolska-Rzechuła 2012: 162, Saraçli i in. 2013: 203).

W celu zbadania wartości CCC dla szesnastu województw reprezentowanych przez frekwencje względne 100 najczęstszych imion w całym zbiorze z lat 1946–2010 (oddzielnie dla imion męskich i żeńskich) stworzono każdorazowo macierz odległości przy pomocy innej metryki odległości, a następnie dla każdej z nich zbudowano dendrogram przy pomocy innej metody wiązania. Co prawda wcześniej dokonałem już wyboru metryki euklidesowej, jednak warto przyjrzeć się, jak „działają” różne metody łączenia skupień także w innych warunkach. W tabelach 44 i 45 zestawiono wyniki tego badania.⁸²

Tab. 44. Wartości współczynnika korelacji kofenetycznej dla wybranych metryk odległości i metod łączenia. Imiona żeńskie (1946-2010)

	ward.D	ward.D2	single	complete	average	mcquitty	median	centroid
euclidean	0,804	0,827	0,854	0,834	0,883	0,851	0,722	0,814
manhattan	0,696	0,724	0,788	0,761	<u>0,827</u>	0,815	0,673	0,704
delta	0,659	0,695	0,776	0,715	<u>0,821</u>	0,805	0,609	0,720
argamon	0,664	0,707	0,801	0,715	<u>0,838</u>	0,825	0,619	0,690
eder	0,643	0,698	0,764	0,719	<u>0,805</u>	0,725	0,563	0,656
simple	0,694	0,724	0,800	0,779	<u>0,841</u>	0,802	0,640	0,748
canberra	0,692	0,725	0,789	0,761	<u>0,845</u>	0,791	0,576	0,756
cosine	0,786	0,786	0,808	0,790	<u>0,822</u>	0,812	0,786	0,804
średnia	0,705	0,736	0,797	0,759	0,835	0,803	0,649	0,737

Tab. 45. Wartości współczynnika korelacji kofenetycznej dla wybranych metryk odległości i metod łączenia. Imiona męskie (1946-2010)

	ward.D	ward.D2	single	complete	average	mcquitty	median	centroid
euclidean	0,810	0,822	0,854	0,817	0,882	0,872	0,716	0,831
manhattan	0,764	0,788	0,804	0,720	<u>0,865</u>	0,800	0,681	0,790
delta	0,698	0,682	0,772	0,713	<u>0,838</u>	0,739	0,596	0,639
argamon	0,652	0,685	0,783	0,722	<u>0,842</u>	0,742	0,593	0,792
eder	0,693	0,722	0,762	0,732	<u>0,840</u>	0,739	0,622	0,799
simple	0,660	0,688	0,813	0,693	<u>0,861</u>	0,792	0,688	0,780
canberra	0,688	0,716	0,823	0,718	<u>0,859</u>	0,811	0,684	0,754
cosine	0,791	0,796	0,810	0,786	<u>0,845</u>	0,831	0,809	0,816
Średnia	0,720	0,737	0,802	0,738	0,854	0,791	0,674	0,775

Jak wynika z danych zamieszczonych w powyższych tabelach, niezależnie od wybranej metryki odległości, najbardziej efektywną metodą łączenia skupień dla materiału imienniczego okazała się metoda średnich połączeń (average). Tak w przypadku imion żeńskich, jak i męskich współczynnik CCC dla tej metody zawsze osiągał wartość powyżej 0,8 i był najwyższy w porównaniu z innymi metodami. Szczególnie istotne jest dla nas to, jak

⁸² Etykiety wierszy stanowią skrótowe nazwy metryk odległości, natomiast etykiety kolumn – skrótowe nazwy metod wiązań, stosowane w pakiecie narzędzi do badań stylometrycznych *stylo* (Eder i in. 2013).

realizują się wartości CCC dla poszczególnych metod wiązania w odniesieniu do macierzy odległości zbudowanej przy pomocy metryki euklidesowej, jednak nie na materiale ogólnym, obejmującym imiona nadawane w poszczególnych województwach w całym okresie 1946-2010, ale na zbiorach dużo mniejszych, odnoszących się do pojedynczych lat z tego przedziału. Jak wiadomo choćby z badań stylometrycznych, wielkość badanego zbioru (np. tekstu) może mieć wpływ na wyniki klasteryzacji. Wykonałem zatem bardziej szczegółowe badanie współczynnika korelacji kofenetycznej dla każdej z dostępnych metod wiązania w odniesieniu do macierzy odległości zbudowanej osobno dla każdego roku. Wyniki tego badania przedstawione zostały w tabeli 46. Zamieszczono w niej kolejno maksymalne i minimalne oraz uśrednione wartości CCC dla każdej z metod wiązania, jakie uzyskałem w toku analizy w oparciu o wyniki wszystkich 65 (dla każdego roku) przeprowadzonych klasteryzacji.

Tab. 46. Wartości maksymalne, minimalne i średnie CCC dla wybranych metod łączenia skupień na bazie macierzy odległości euklidesowych uzyskane na podstawie 65 klasteryzacji (dla kolejnych lat z przedziału 1946–2010)

	ward.D	ward.D2	single	complete	average	mcquitty	median	centroid
maks.								
żeńskie	0,885	0,886	0,894	0,890	<u>0,904</u>	0,901	0,878	0,873
męskie	0,887	0,896	0,900	0,903	<u>0,923</u>	0,915	0,888	0,902
min.								
żeńskie	0,509	0,558	0,542	0,591	<u>0,727</u>	0,613	-0,017	0,348
męskie	0,542	0,565	0,540	0,536	<u>0,647</u>	0,595	0,352	0,537
średnia								
żeńskie	0,729	0,746	0,761	0,772	<u>0,825</u>	0,807	0,622	0,725
męskie	0,705	0,728	0,745	0,743	<u>0,806</u>	0,786	0,646	0,763

Podobnie jak w poprzednim badaniu, również w przypadku klasteryzacji obiektów o mniejszej wielkości (zawierających mniejsze zbiory imion) najwyższe wartości CCC odnotowano dla metody średnich połączeń (average). Co prawda, w przypadku niektórych lat wartość ta wynosiła mniej niż postulowane 0,75 (lub 0,8), jednak w porównaniu z innymi metodami wiązania wartości minimalne CCC były zdecydowanie wyższe dla tej metody. Bazując na tych obserwacjach, a także na fakcie osiągnięcia przez tę metodę najwyższej wartości średniej współczynnika CCC, postanowiłem zastosować ją do zbudowania dendrogramów, na podstawie których w następującym podrozdziale V.5 będziemy mogli prześledzić konfiguracje przestrzenne badanych przez mnie obiektów oraz ich zmienność w czasie.

V.5. Wyniki hierarchicznej analizy skupień – imiennictwo na obszarze współczesnych województw (lata 1946–2010)

Na rys. 85 i 86 przedstawione zostały wyniki hierarchicznej analizy skupień przeprowadzonej (według opisanej wcześniej metody) na bazie imiennictwa z obszaru 16 współczesnych województw w każdym z lat w przedziale 1946–2010 (oddzielnie dla imiennictwa żeńskiego i męskiego)⁸³. Każdy z rysunków składa się z 65 pięciu paneli (dla kolejnych lat) złożonych z dwóch elementów – dendrogramu i mapy. Najistotniejszym elementem jest dla nas sam dendrogram, na którym każdy z „liści” odpowiada danemu województwu oznaczonemu skrótem. Na jego podstawie możemy ocenić siłę powiązań pomiędzy poszczególnymi województwami oraz to, w jakiej kolejności łączyły się ze sobą w kolejne zgrupowania. Dodatkowo szarą przerywaną linią na dendrogramach zaznaczono cztery największe skupienia, które otrzymaliśmy w toku analizy. Czarną przerywaną pionową linią oznaczono wysokość dendrogramu (h) potrzebną do uzyskania czterech skupień (im dłuższa gałąź dendrogramu – tym zróżnicowanie między obiektami jest większe).

Drugi z elementów – mapa – to kartograficzny obraz tych czterech największych skupień. Województwa, które wspólnie zostały zaklasyfikowane do jednego z nich, zostały oznaczone tym samym odcieniem. Liczba czterech klastrów została przeze mnie wybrana arbitralnie. Miało to na celu sprawdzenie, czy na przestrzeni całego badanego okresu dominował jakiś określony typ regionalizacji, czy też układ danych ulegał przeobrażeniom. Pozwoliło to też na uwypuklenie najbardziej charakterystycznych zjawisk.

Analizując wyniki klasteryzacji na dendrogramach z rys. 85 i 86, możemy zaobserwować, że w zakresie układu imion na listach frekwencyjnych wyodrębniają się pewne bardzo wyraźne geograficzne skupiska, obejmujące sąsiadujące ze sobą województwa, których obszar często pokrywa się z określonymi regionami historyczno-kulturowymi.

Tak w przypadku imion żeńskich, jak i męskich województwa małopolskie i podkarpackie (pokrywające się w ogromnej większości z obszarem dawnej Galicji) na przestrzeni całego badanego okresu stanowiły zawsze odrębne skupienie (zawsze były dla siebie parą w skupieniu drugiego rzędu). Ponadto do roku 1980 w przypadku imion męskich oraz do roku 1974 w przypadku imion żeńskich (a także i w niektórych późniejszych latach) oba te województwa tworzyły jedno z czterech największych skupień, następnie łącząc się w jednym klastrze z sąsiednim województwem lubelskim i/lub świętokrzyskim, które składają się w dużej części na historyczną Małopolskę.

⁸³ Z uwagi na obszerność rysunków prezentuję je dopiero po części tekstowej niniejszego podrozdziału (od s. 193).

Podobnie wygląda to w przypadku województw w dużej części pokrywających się z historycznym Górnym Śląskiem, czyli opolskim i śląskim (obejmującym pogranicze śląsko-małopolskie). Generalnie w większości przypadków województwa te wspólnie tworzą jedno skupienie, choć w poszczególnych latach woj. opolskie odróżnia się na tyle od woj. śląskiego, jak i pozostałych województw, że tworzy samo w sobie odrębny klaster na poziomie czterech największych skupień, co jest widoczne zwłaszcza w przypadku imion męskich (1947, 1958–1971, 1976, 1978).

Innym przykładem wyraźnej odrębności tylko jednego obiektu, jak w przypadku województwa opolskiego, jest wspomniane już województwo podlaskie, które zwykle wyraźnie odstawało od pozostałych zgrupowań, często tworząc samodzielnie jedno z czterech największych skupień. W innych wypadkach łączyło się w jednym skupieniu z sąsiednim woj. warmińsko-mazurskim lub lubelskim (w zakresie imiennictwa żeńskiego cezurą dla tych dwóch połączeń są lata 70., natomiast w przypadku imion męskich nie jest to tak wyraźnie widoczne).

Ta często obserwowana wyraźna odrębność województw opolskiego i podlaskiego wiąże się zapewne z faktem, iż są to tereny najliczniej zamieszkane przez przedstawicieli mniejszości narodowych – niemieckiej na Opolszczyźnie oraz białoruskiej i litewskiej na Podlasiu. Częstsze nadawane na ich obszarze imion rzadziej spotykanych gdzie indziej, swego rodzaju antroponimicznych regionalizmów frekwencyjnych, nawet jeśli nie są to imiona najpopularniejsze, osłabia frekwencję względną innych imion, co ma wpływ na odmienny układ list frekwencyjnych niż na pozostałych obszarach.⁸⁴

Do nieco mniej regularnych, ale jednak często tworzących jedno z czterech największych klastrów regionalnych możemy zaliczyć także połączenie województw mazowieckiego i łódzkiego (ze sporadycznym udziałem świętokrzyskiego) oraz lubelskiego, a więc obszarów przed 1918 r. należących do Królestwa Kongresowego oraz (zwłaszcza w przypadku imion żeńskich) województw kujawsko-pomorskiego, pomorskiego i wielkopolskiego (do których sporadycznie dołącza też opolskie i/lub śląskie), w większej części związanych z państwem pruskim sprzed 1918 r.

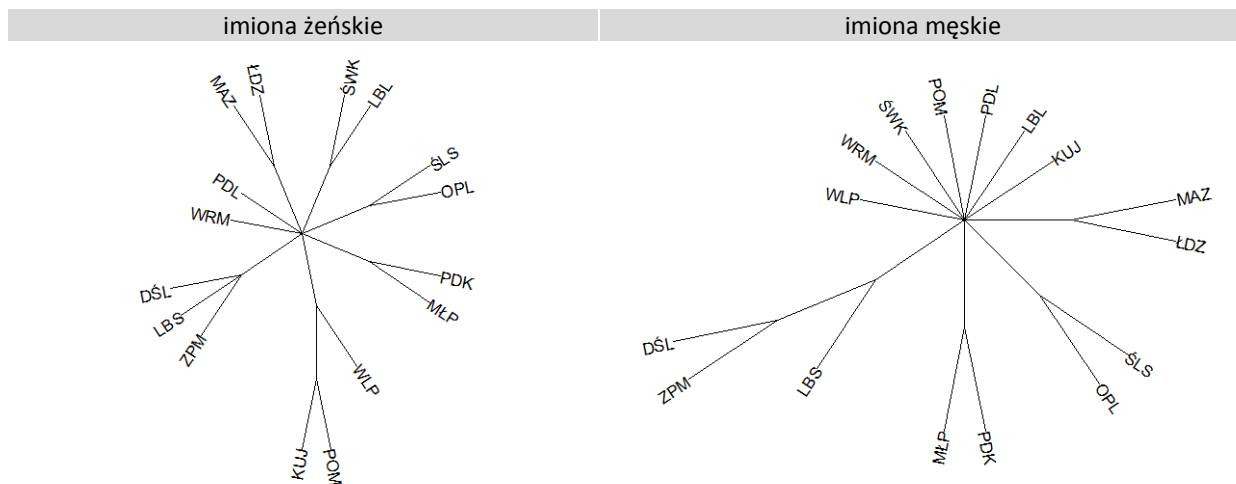
Województwa tworzące ścianę zachodnią Polski, czyli dolnośląskie, lubuskie i zachodniopomorskie, choć nie tworzą nigdy jednego z czterech największych klastrów, to

⁸⁴ Przeprowadzone przeze mnie wcześniej badania regionalnej specyfiki w nadawaniu imion w latach 1995–2010 potwierdziły, że w najczęstszym imiennictwie woj. opolskiego występowała znaczna liczba imion niemieckich lub popularnych na niemieckim obszarze językowym, które są rzadziej spotykane w antroponimii innych województw w Polsce (Swoboda 2013: 57).

niemal bez wyjątku tworzą odrębne skupienie. Jest to ciekawe zjawisko, zważywszy na fakt, iż w odróżnieniu od Górnego Śląska, czy historycznej Galicji, cechujących się dużym odsetkiem ludności autochtonicznej, na obszarach tych po 1945 r. osiedlała się ludność rozmaitej proveniencji regionalnej. Pomimo tego (lub może właśnie dlatego?) imiennictwo tych obszarów jest ze sobą silnie powiązane.

Opisane powyżej najbardziej charakterystyczne skupienia dobrze uwidaczniają się w swego rodzaju zbiorczej analizie dendrogramów, wykonanej przy użyciu metody tzw. drzewek konsensusu (ang. *bootstrap consensus tree*, w skrócie BCT, zob. Eder 2013, Eder 2014: 99). Ma ona na celu oszacowanie dendrogramu uśrednionego na podstawie wielu tradycyjnych dendrogramów (takich, jakie wygenerowałem dla poszczególnych lat) – poszukiwane są najbardziej stabilne gałęzie drzewek, pojawiające się najczęściej na każdym z dendrogramów i na ich podstawie rekonstruuje najsilniejsze podobieństwa między poszczególnymi obiektami (Eder 2014: 99), w tym przypadku zbiorami imion. Efekty analizy dendrogramów metodą konsensusową widoczne są na rys. 87. Warto zauważyć, że w obrębie imiennictwa żeńskiego wykrystalizowały się dużo wyraźniejsze podziały regionalne, obejmujące wszystkie wymienione wcześniej zgrupowania, natomiast w przypadku imion męskich widoczne są tylko niektóre z nich.

Rys. 87. Analiza BCT wykonana na podstawie 65 dendrogramów dla kolejnych lat z przedziału 1946–2010 (zob. rys. 82 i 83)



Jeśli chodzi o rozmieszczenie geograficzne czterech największych skupień (nieco odmienne w imiennictwie żeńskim i męskim) ukazane na kartogramach, to nie możemy powiedzieć, że ma ono charakter trwały. Na przestrzeni całego okresu ich konfiguracja ulegała licznym przeobrażeniom, na co jednak mogły mieć wpływ znacząco odstające na tle

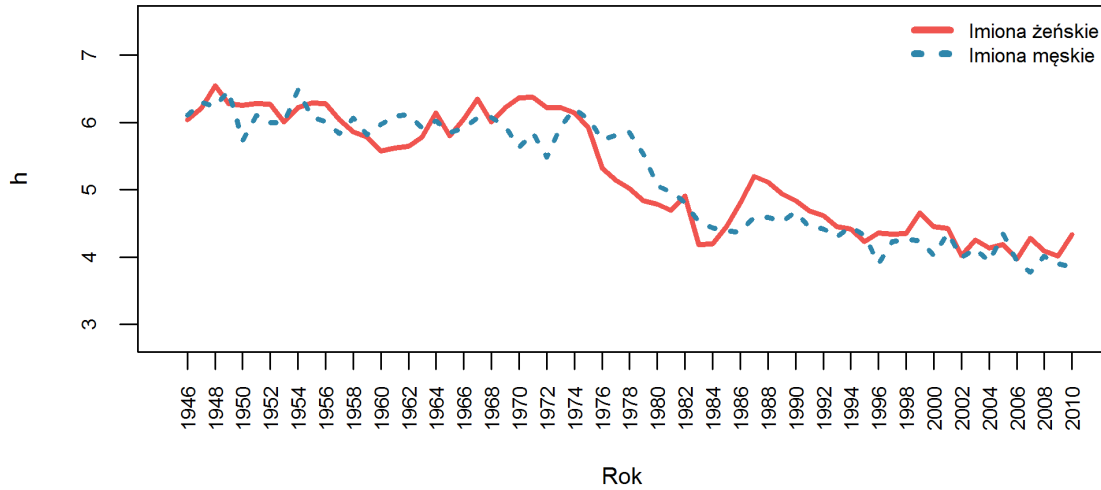
innych obszarów województwa opolskie i podlaskie. W przypadku imion żeńskich w latach 1946–54, 1965–74 czy 1988–94 granice największych skupień na ogół pokrywały się z granicami dawnych państw zaborczych⁸⁵ (oczywiście woj. podlaskie zwykle tworzyło odrębne skupienie, ale na dendrogramie w dalszej kolejności łączyło się z pozostałym obszarem Kongresówki). W innych latach taki wyraźny podział ulegał większemu lub mniejszemu zatarciu – ujawniały się częściej skupienia obejmujące np. wschodnią lub południowo-wschodnią część kraju (bez wyróżnienia np. Galicji). Warto też zauważyć, że w pierwszych latach badanego okresu obszar Ziem Zachodnich tworzył wspólny klaster z obszarem dawnych Prus, później zaś częściej wykazywał więcej podobieństw do wschodniej części Polski. W imiennictwie męskim do końca lat 70. mogliśmy obserwować w miarę stabilny układ czterech największych klastrów uwzględniających podział na Galicję, Górny Śląsk, Podlasie (czasem też cały pas województw wschodnich – lubelskie, podlaskie i warmińsko-mazurskie) oraz pozostały obszar Polski. Natomiast od lat 80. układ głównych skupień ulega ciągłym rekonfiguracjom.

Jakkolwiek na przestrzeni całego badanego okresu, tj. lat 1946–2010, uwidaczniają się wyraźne skupienia województw różniących się pod kątem układu imion na listach frekwencyjnych, to różnice te z upływem czasu ulegały zmniejszeniu. Obserwując na dendrogramach dla następujących po sobie lat wartość wysokości odcięcia czterech największych skupień (pionowa linia przerywana), zauważymy, że odległość między poszczególnymi klastrami obiektów od lat 70. systematycznie malała i to zarówno w obrębie imiennictwa męskiego, jak i żeńskiego. Wartości te obrazują krzywe na wykresie z rys. 85. Początkowo do uzyskania czterech skupień musieliśmy „obciąć” dendrogram na wysokości między 6 a 7, natomiast w ostatnich latach był to poziom rzędu 4–5, co jest tożsame ze spadkiem zróżnicowania pomiędzy tymi skupieniami. Jeszcze lepiej zjawisko to oddaje wykres na rys. 86, na którym przebieg krzywych obrazuje wysokość całego dendrogramu, tj. wysokość na jakiej wszystkie obiekty tworzą jedno skupienie. Jak łatwo zauważyć, na przestrzeni całego badanego okresu zmalała ona z ok. 13 (imiona żeńskie) lub ok. 10 (imiona męskie) do 5–6. Zatem choć różnice w obrębie najczęstszego imiennictwa między

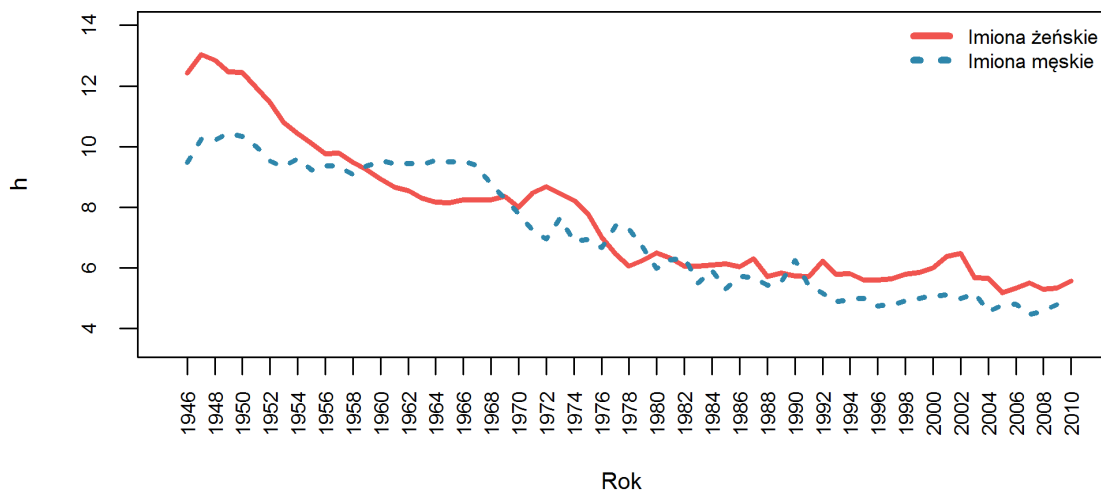
⁸⁵ Mam oczywiście na uwagę fakt, że granice współczesnych województw nie pokrywają się w pełni z granicami dawnych Prus, Austro-Węgier i Cesarstwa Rosyjskiego. W celu weryfikacji, czy rzeczywiście możemy mówić o śladach granic zaborowych przejawiających się w imiennictwie, w rozdziale V.6 dokonam nieco bardziej szczegółowej analizy na materiale imion nadawanych w powiatach województwa śląskiego, którego obszar w przeszłości był podzielony pomiędzy wspomniane mocarstwa.

poszczególnymi województwami są nadal widoczne, to na przestrzeni całego badanego okresu doszło do ich znacznej niwelacji.

Rys. 85. Wysokość (h), na której zawiązują się cztery największe skupienia na dendrogramach dla kolejnych lat z przedziału 1946–2010 (zob. rys. 85 i 86)

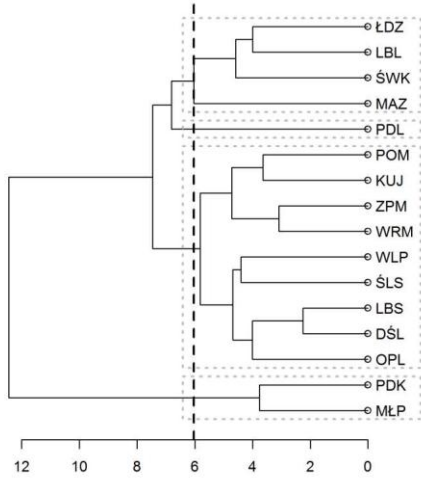


Rys. 86. Wysokość (h), na której zawiązuje się jedno (główne) skupienie na dendrogramach dla kolejnych lat z przedziału 1946–2010 (zob. rys. 85 i 86)

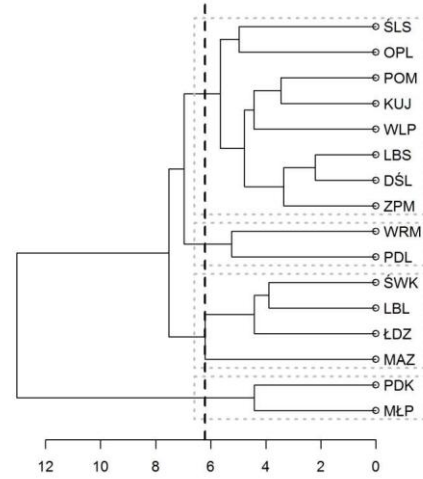


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

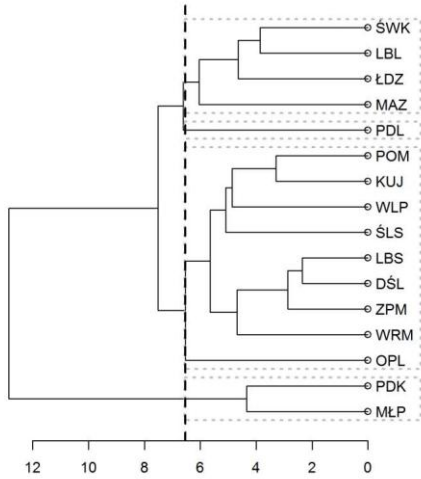
Im. żeńskie, 1946 r., h = 6.04



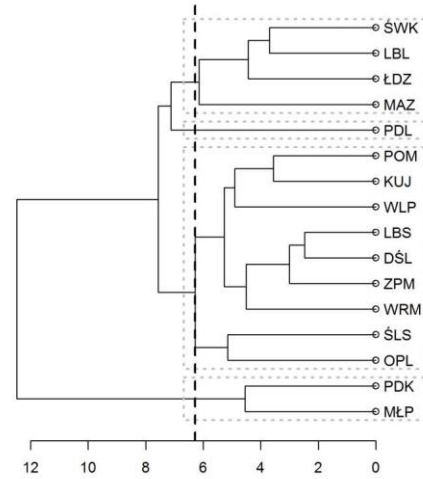
Im. żeńskie, 1947 r., h = 6.21



Im. żeńskie, 1948 r., h = 6.54

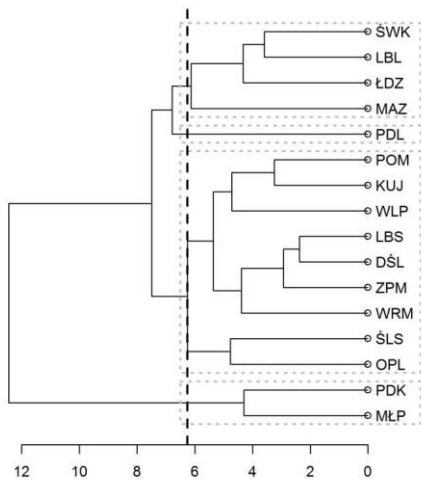


Im. żeńskie, 1949 r., h = 6.28

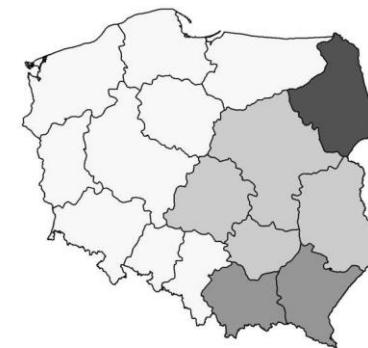
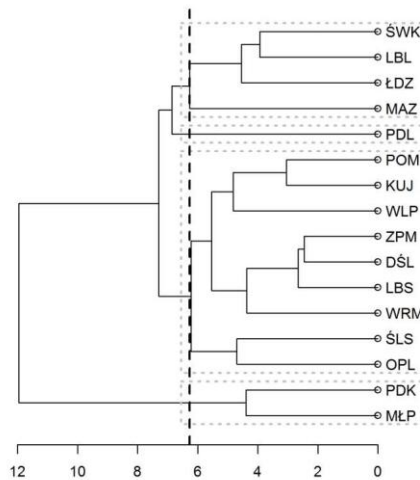


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

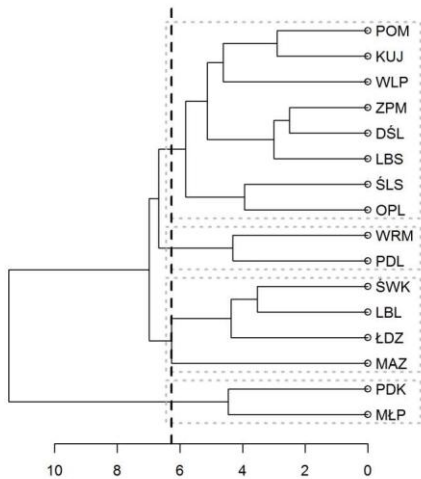
Im. żeńskie, 1950 r., h = 6.25



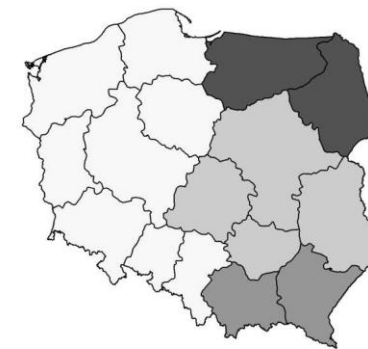
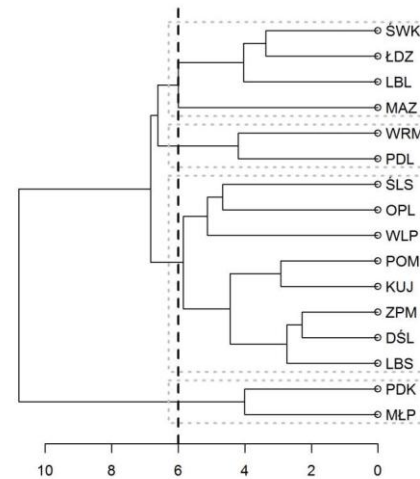
Im. żeńskie, 1951 r., h = 6.27



Im. żeńskie, 1952 r., h = 6.27

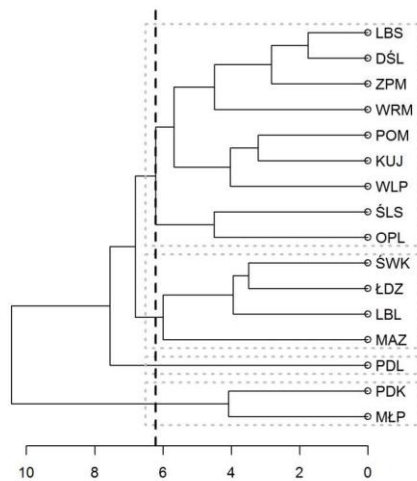


Im. żeńskie, 1953 r., h = 6

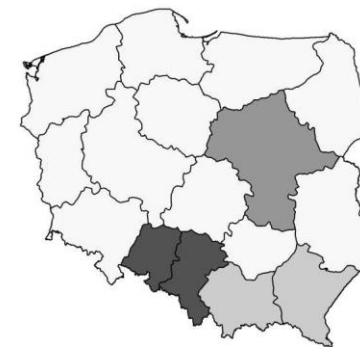
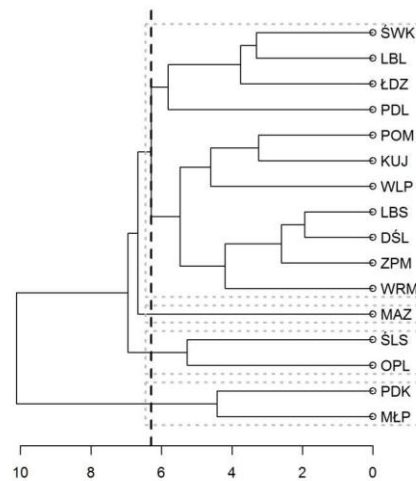


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

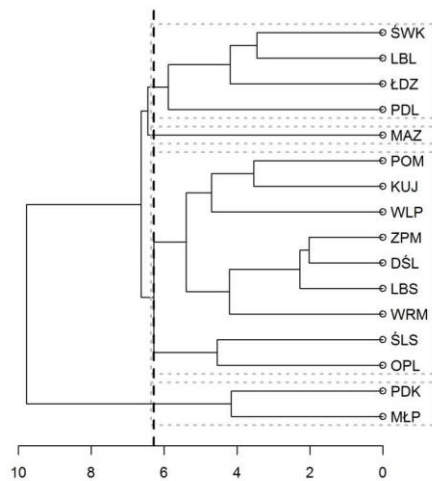
Im. żeńskie, 1954 r., h = 6.22



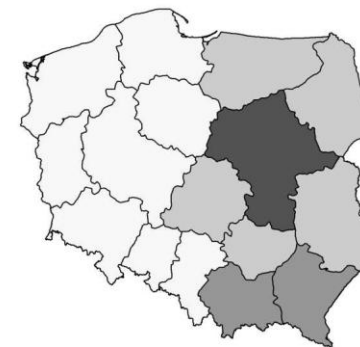
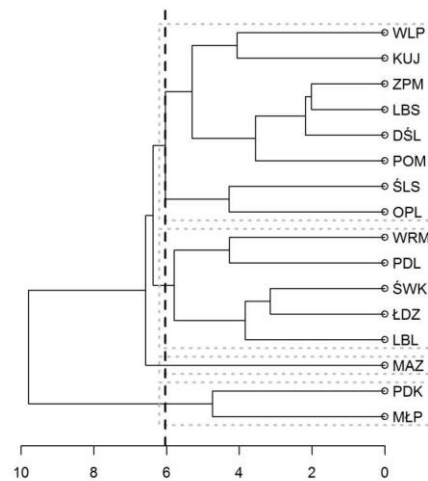
Im. żeńskie, 1955 r., h = 6.29



Im. żeńskie, 1956 r., h = 6.28

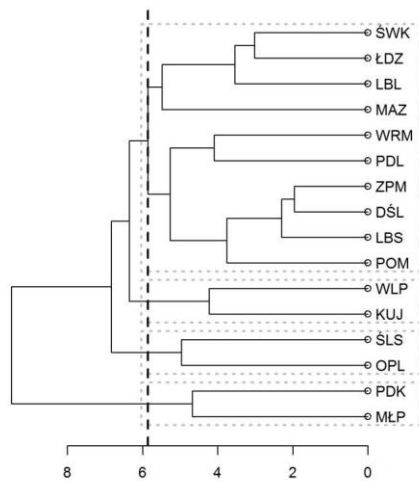


Im. żeńskie, 1957 r., h = 6.04

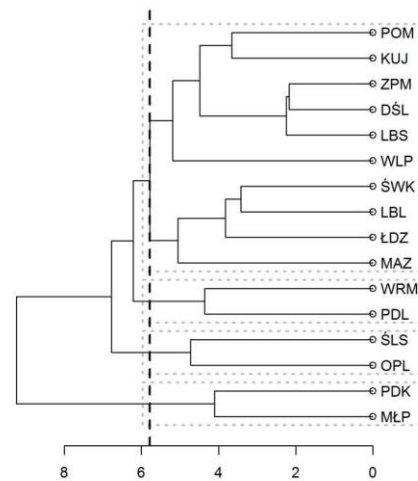


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

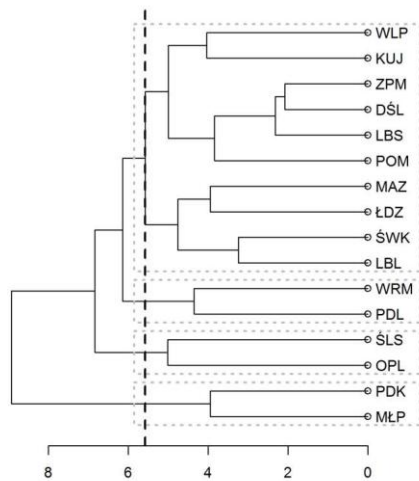
Im. żeńskie, 1958 r., h = 5.86



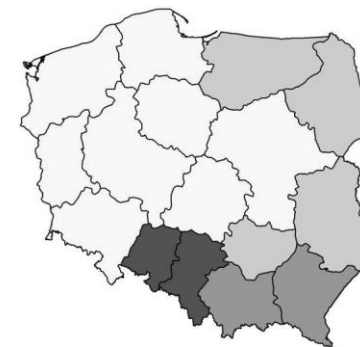
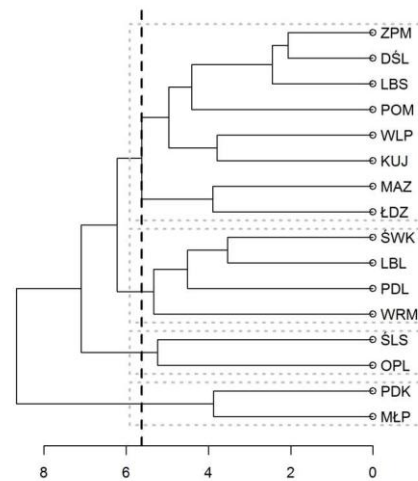
Im. żeńskie, 1959 r., h = 5.78



Im. żeńskie, 1960 r., h = 5.57

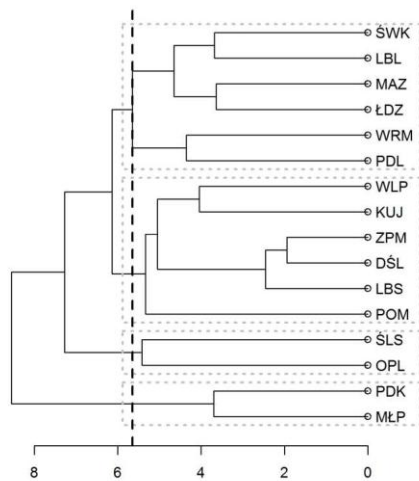


Im. żeńskie, 1961 r., h = 5.62

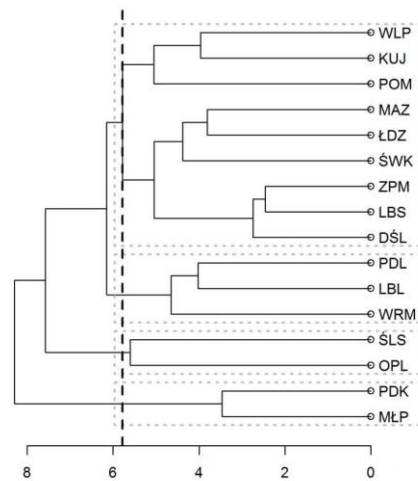


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

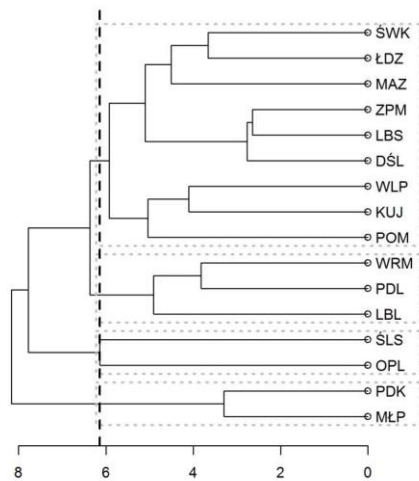
Im. żeńskie, 1962 r., h = 5.65



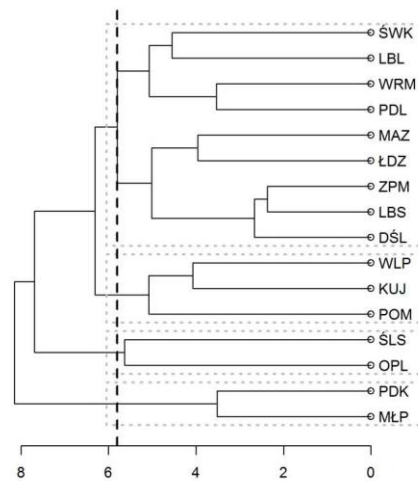
Im. żeńskie, 1963 r., h = 5.78



Im. żeńskie, 1964 r., h = 6.14

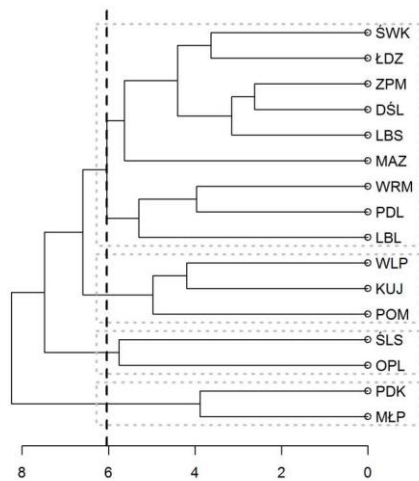


Im. żeńskie, 1965 r., h = 5.8

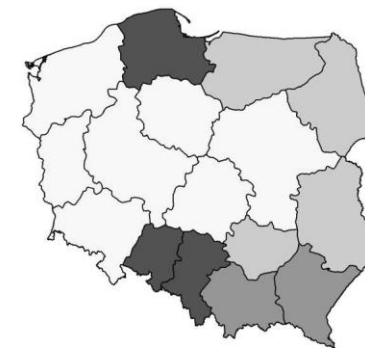
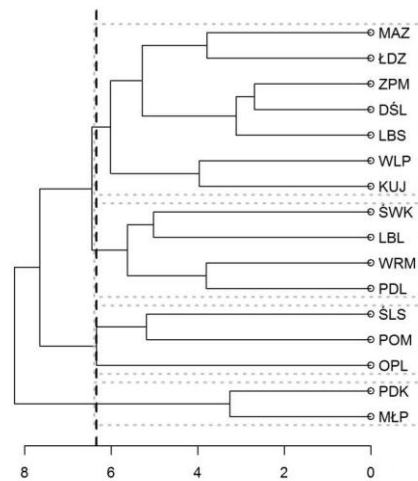


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

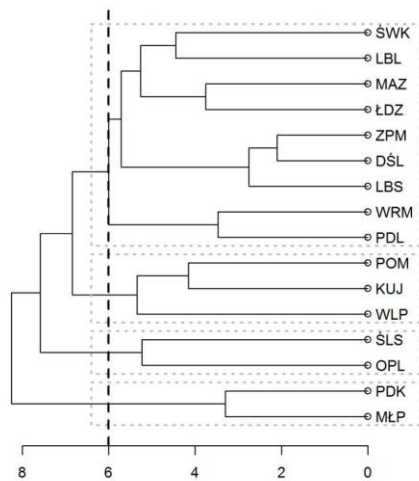
Im. żeńskie, 1966 r., h = 6.05



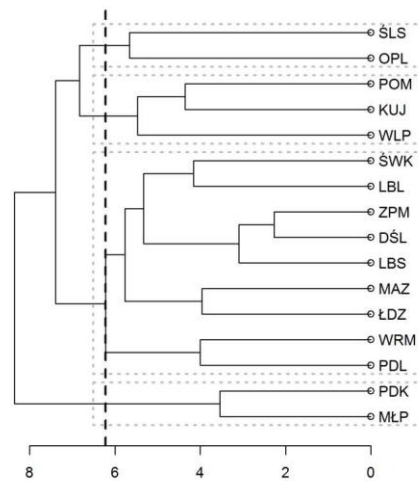
Im. żeńskie, 1967 r., h = 6.35



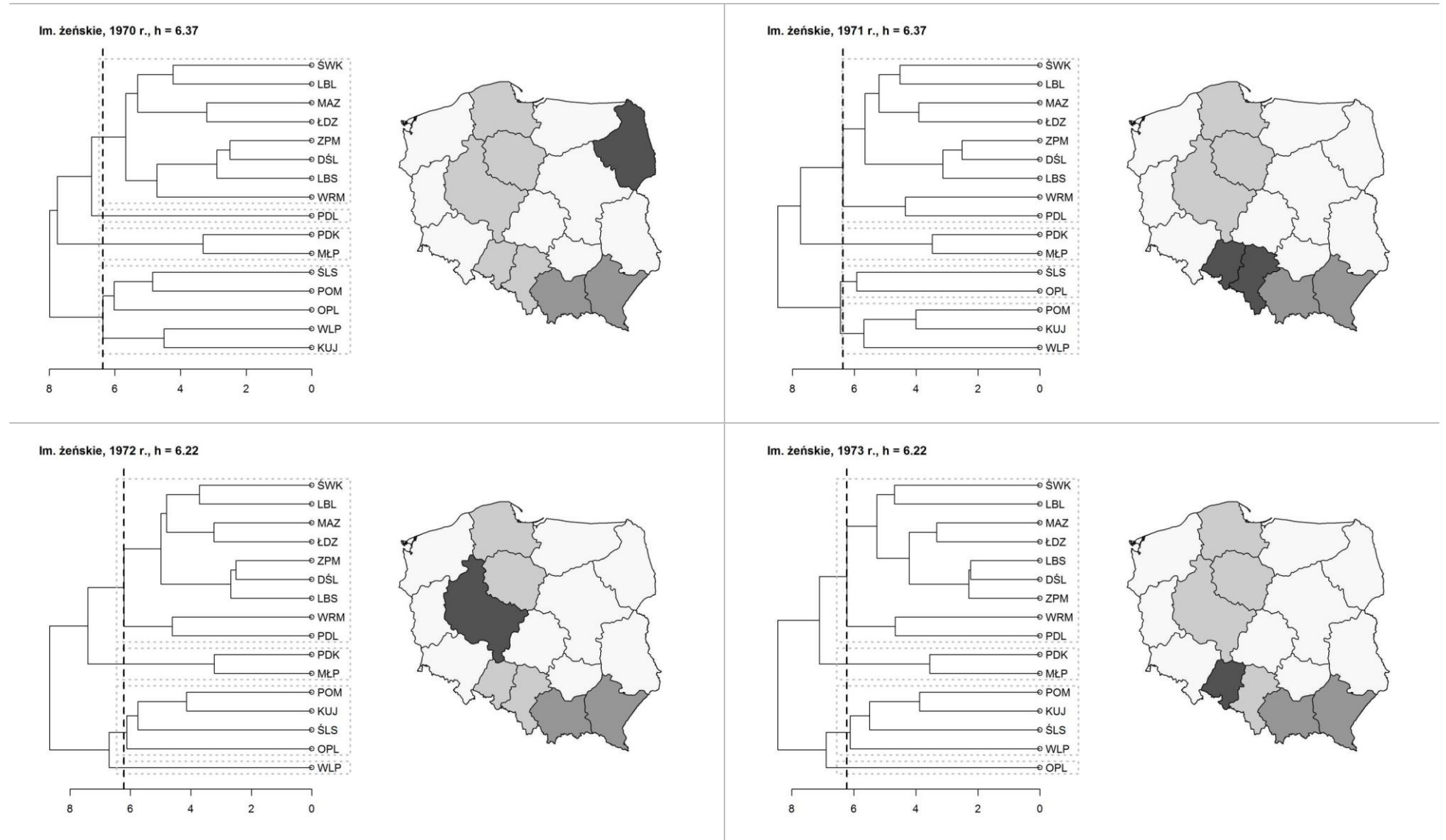
Im. żeńskie, 1968 r., h = 6



Im. żeńskie, 1969 r., h = 6.22

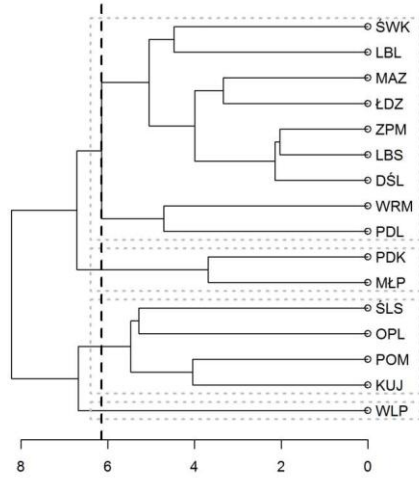


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

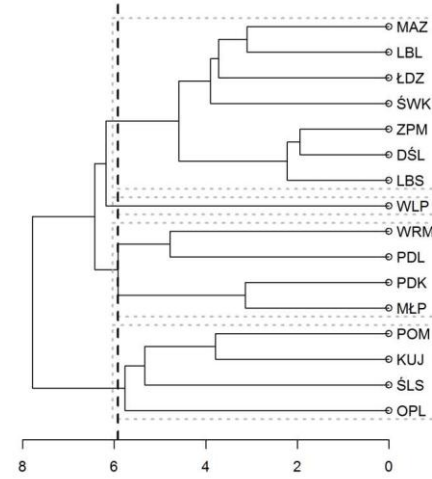


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

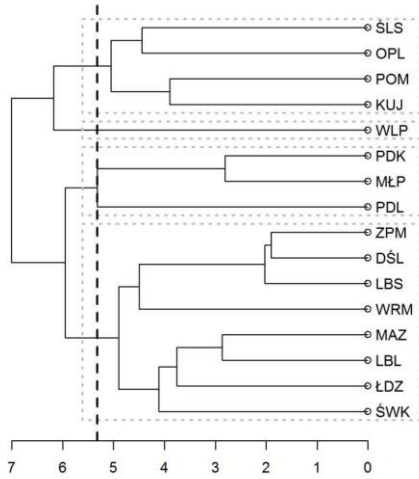
Im. żeńskie, 1974 r., h = 6.14



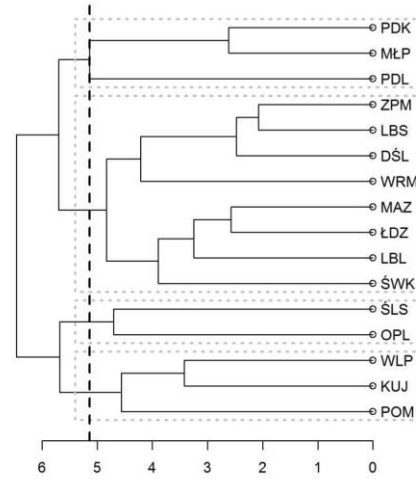
Im. żeńskie, 1975 r., h = 5.92



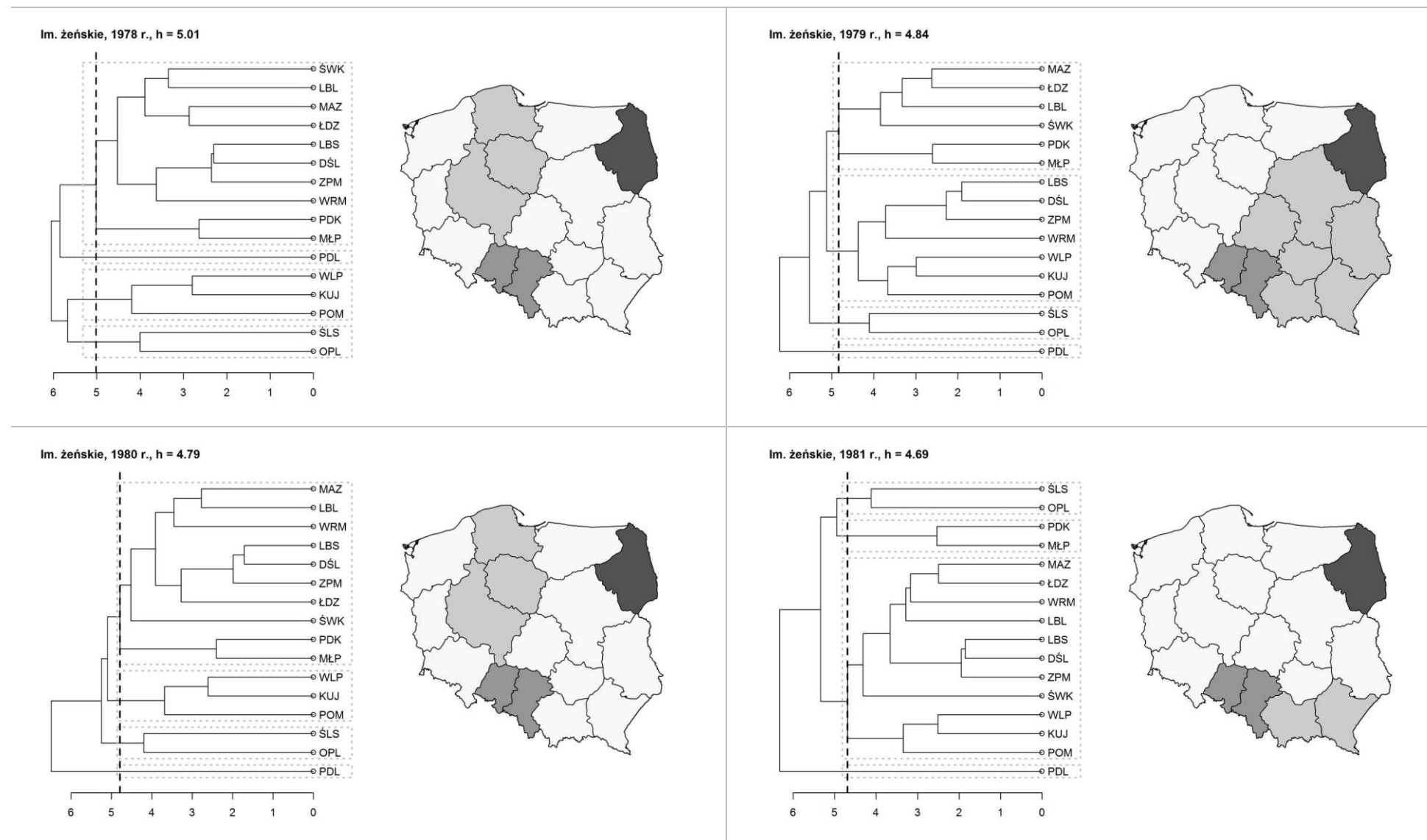
Im. żeńskie, 1976 r., h = 5.32



Im. żeńskie, 1977 r., h = 5.14

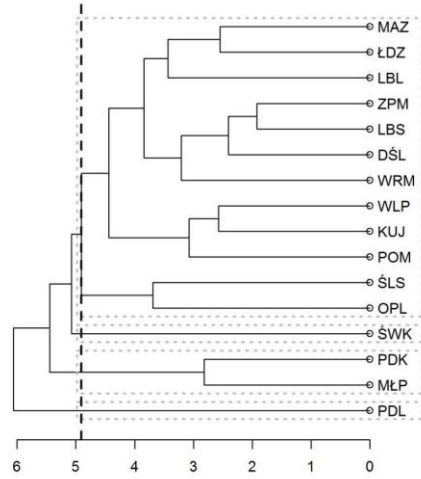


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

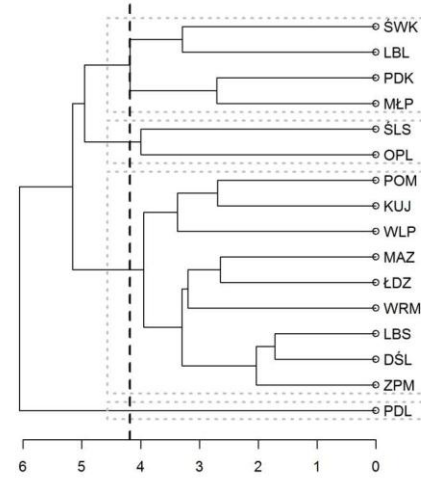


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

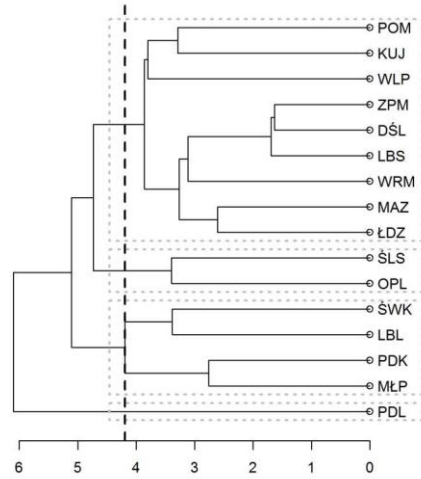
Im. żeńskie, 1982 r., $h = 4.9$



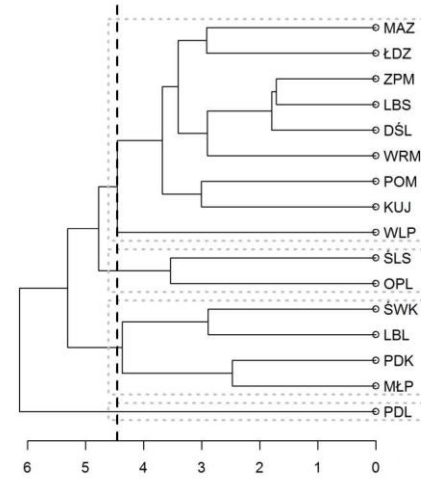
Im. żeńskie, 1983 r., $h = 4.18$



Im. żeńskie, 1984 r., $h = 4.19$

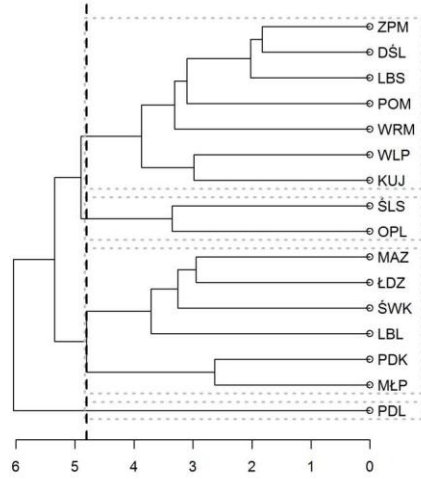


Im. żeńskie, 1985 r., $h = 4.45$

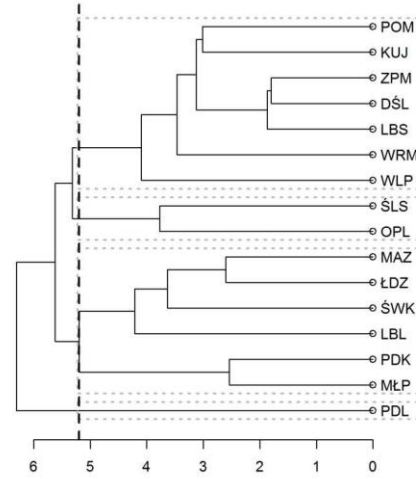


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

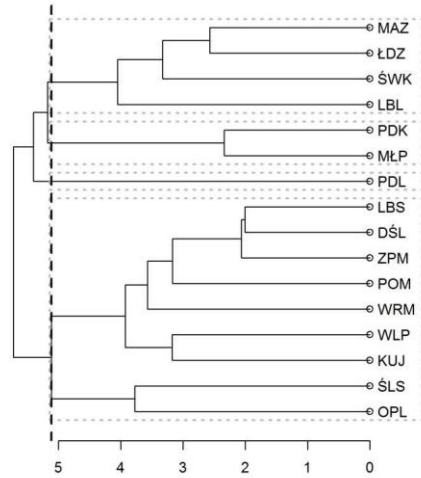
Im. żeńskie, 1986 r., h = 4.8



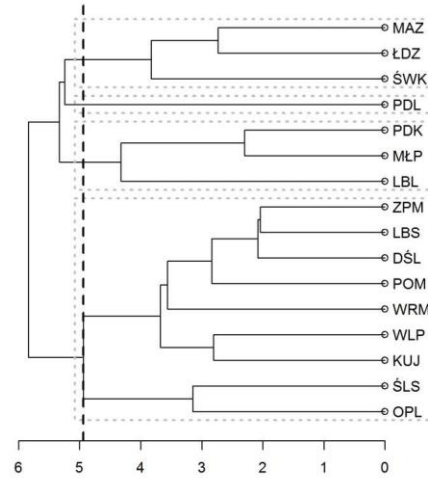
Im. żeńskie, 1987 r., h = 5.2



Im. żeńskie, 1988 r., h = 5.12

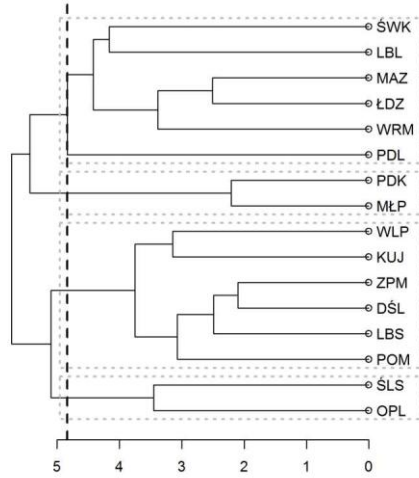


Im. żeńskie, 1989 r., h = 4.94

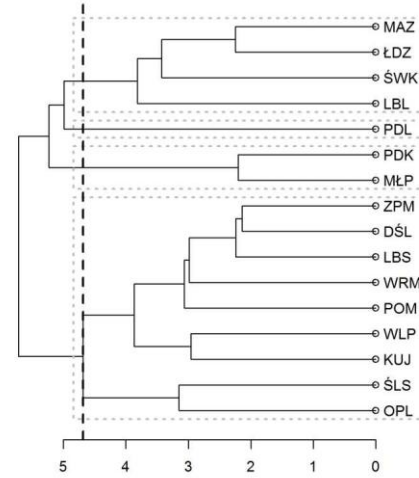


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

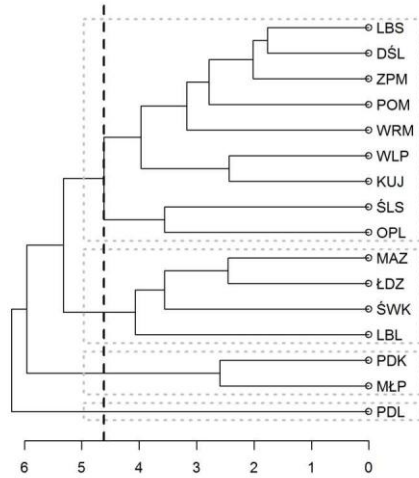
Im. żeńskie, 1990 r., $h = 4.84$



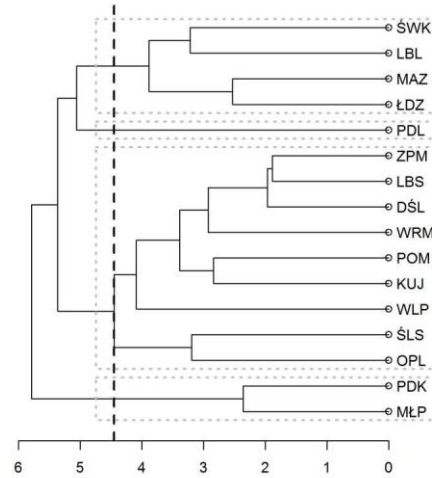
Im. żeńskie, 1991 r., $h = 4.69$



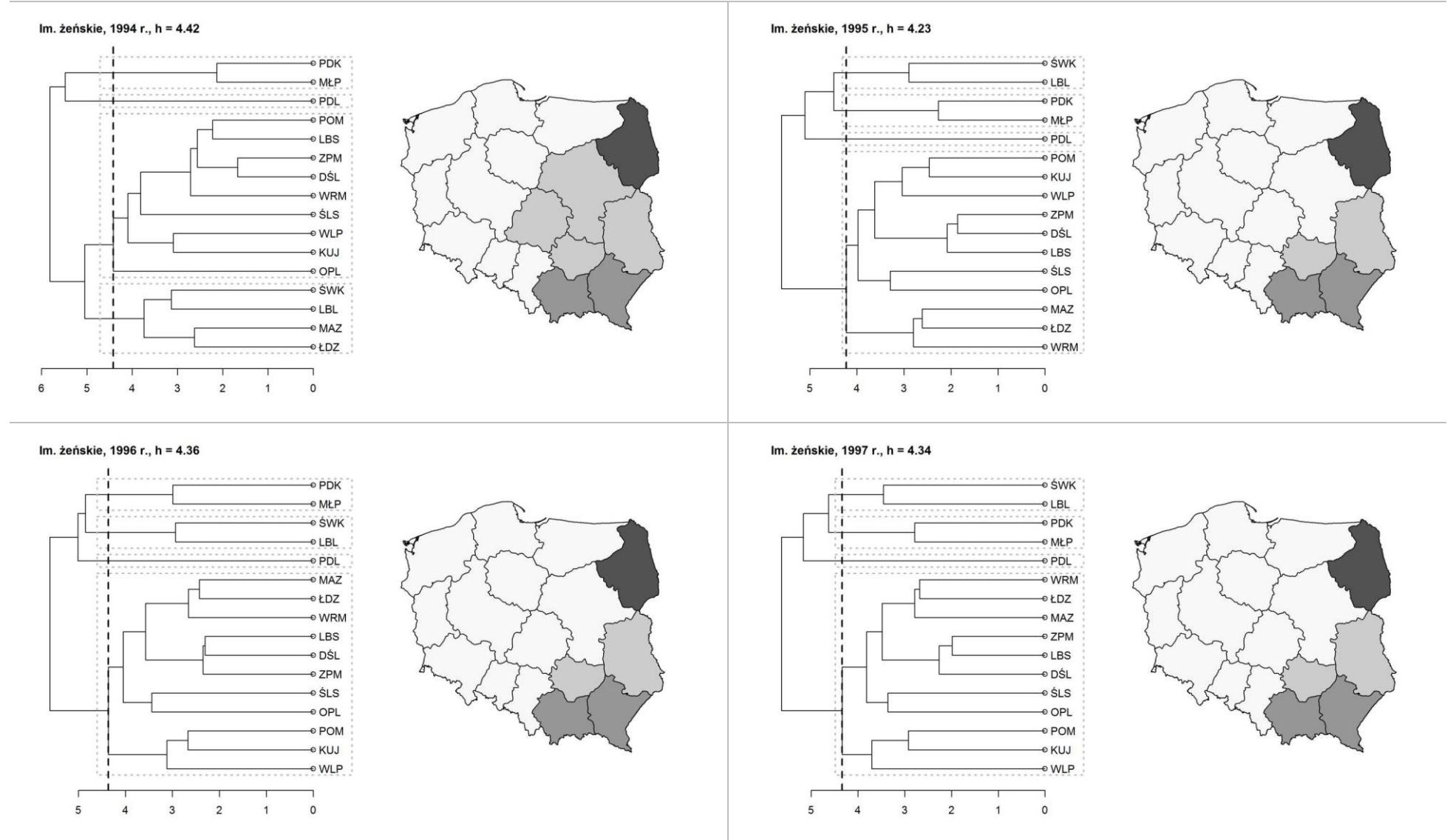
Im. żeńskie, 1992 r., $h = 4.61$



Im. żeńskie, 1993 r., $h = 4.45$

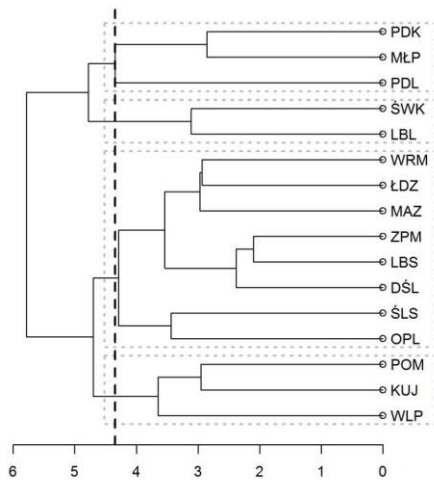


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

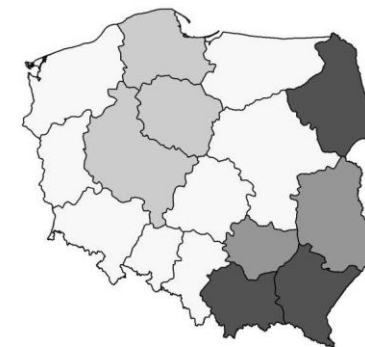
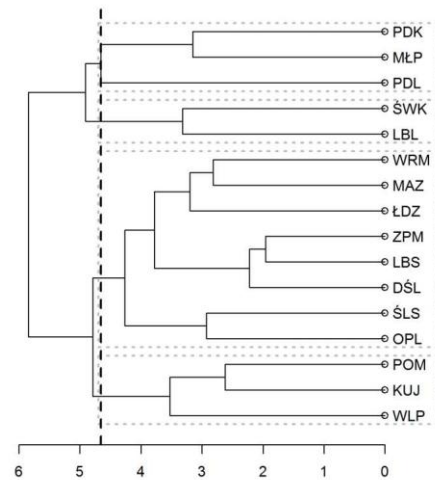


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

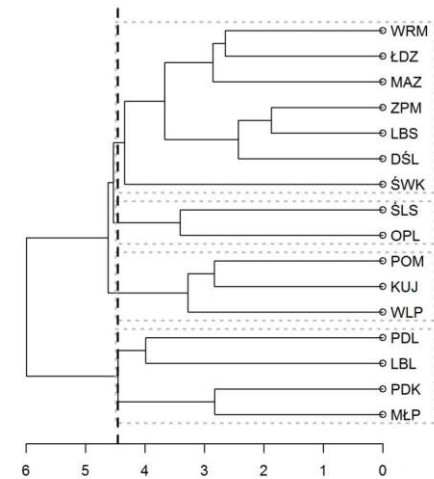
Im. żeńskie, 1998 r., $h = 4.35$



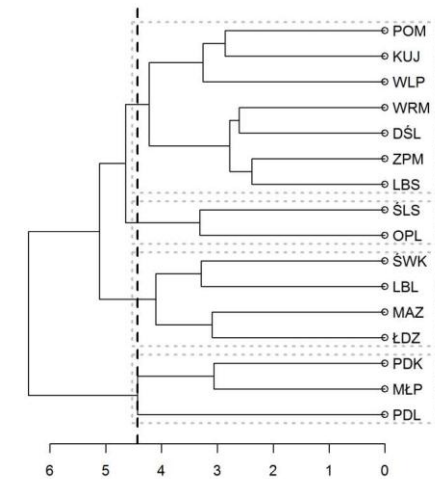
Im. żeńskie, 1999 r., $h = 4.66$



Im. żeńskie, 2000 r., $h = 4.46$

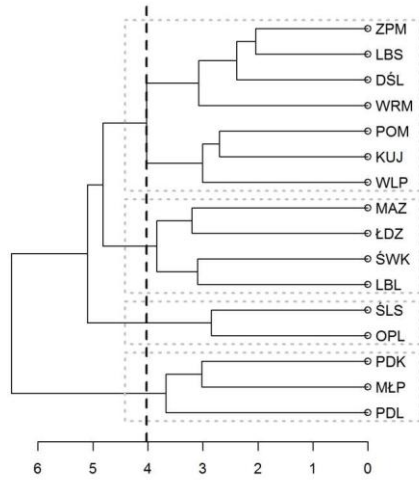


Im. żeńskie, 2001 r., $h = 4.43$

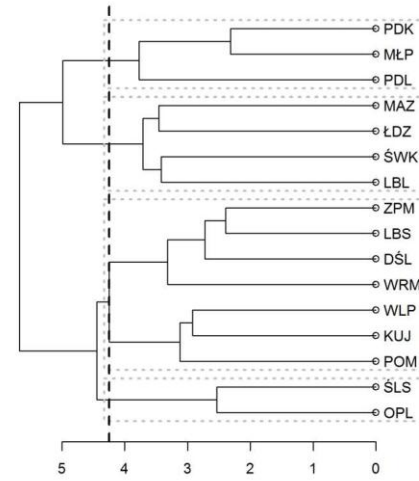


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

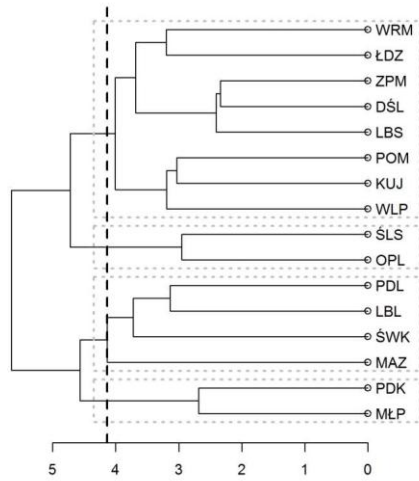
Im. żeńskie, 2002 r., $h = 4.02$



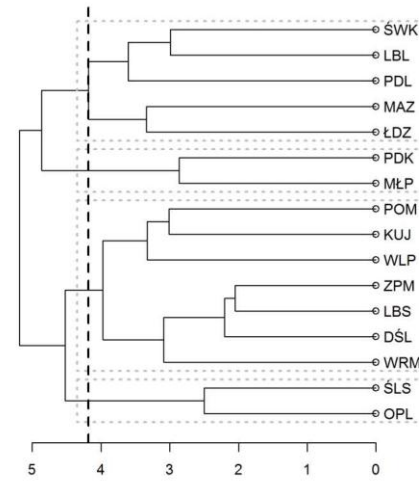
Im. żeńskie, 2003 r., $h = 4.25$



Im. żeńskie, 2004 r., $h = 4.14$

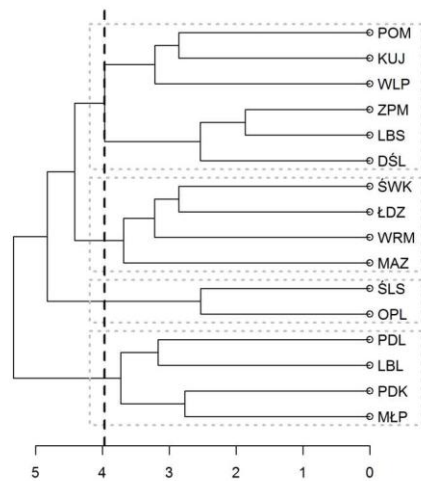


Im. żeńskie, 2005 r., $h = 4.18$

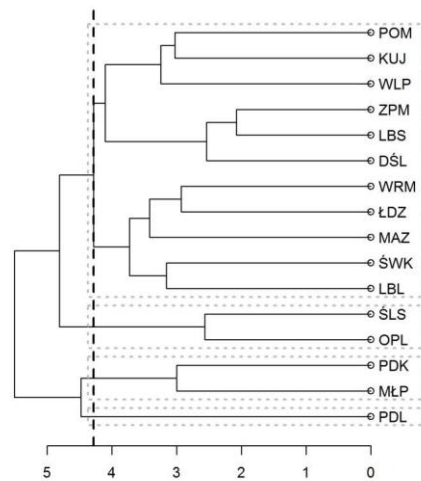


Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

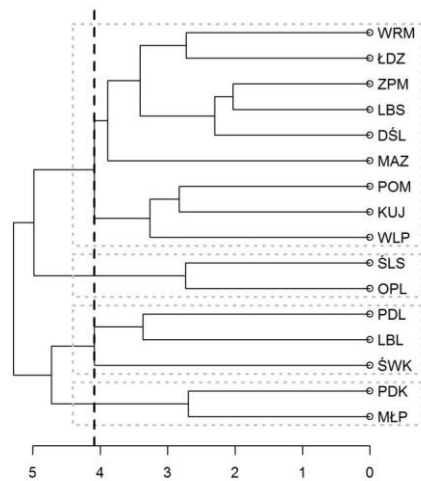
Im. żeńskie, 2006 r., $h = 3.97$



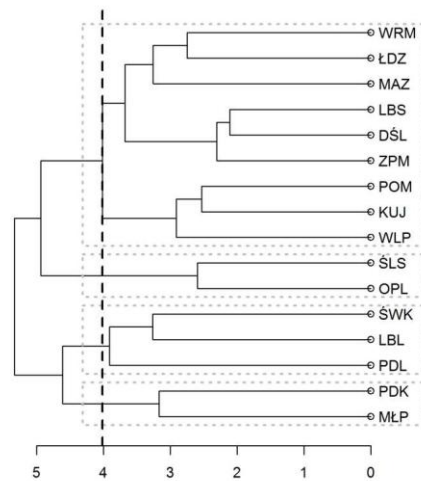
Im. żeńskie, 2007 r., $h = 4.28$



Im. żeńskie, 2008 r., $h = 4.09$

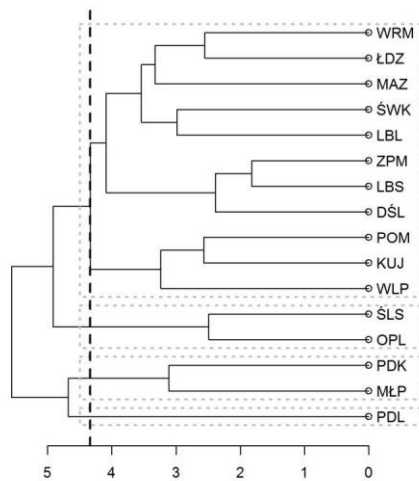


Im. żeńskie, 2009 r., $h = 4.02$



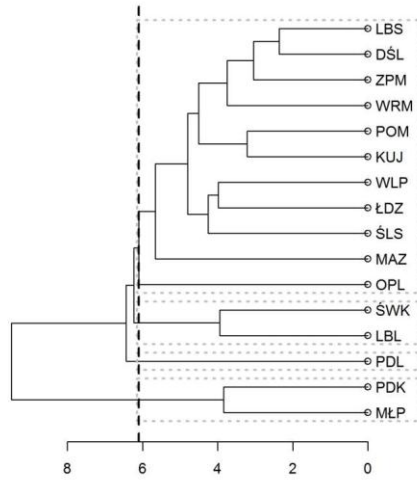
Rys. 85 Hierarchiczna analiza skupień – 100 najczęstszych imion żeńskich. Rozmieszczenie geograficzne czterech największych skupień

Im. żeńskie, 2010 r., $h = 4.33$

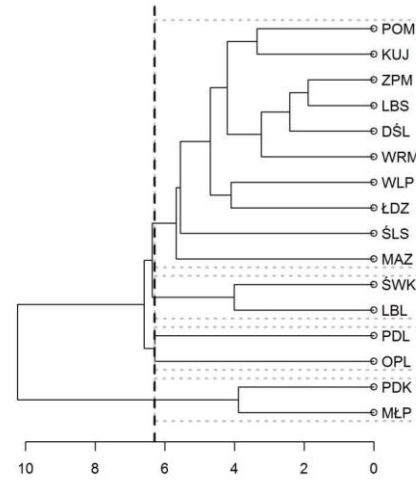


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

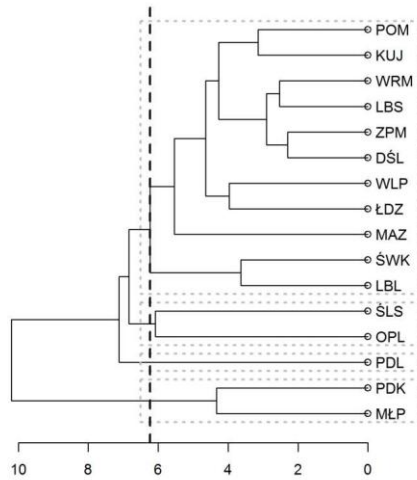
Im. męskie, 1946 r., $h = 6.11$



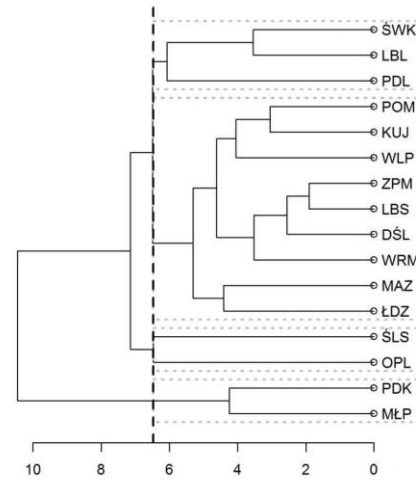
Im. męskie, 1947 r., $h = 6.29$



Im. męskie, 1948 r., $h = 6.23$

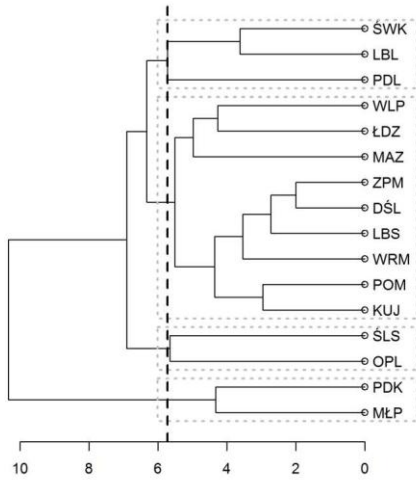


Im. męskie, 1949 r., $h = 6.47$

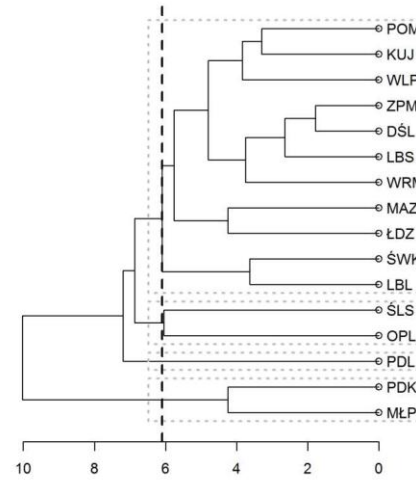


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

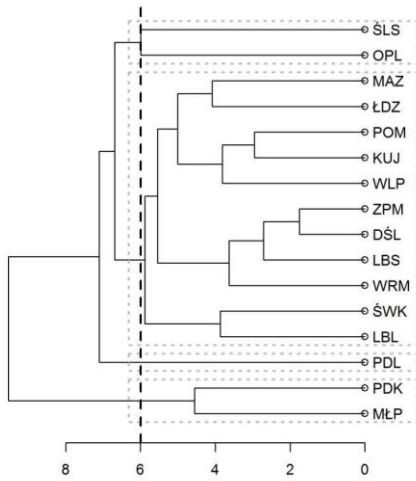
Im. męskie, 1950 r., $h = 5.73$



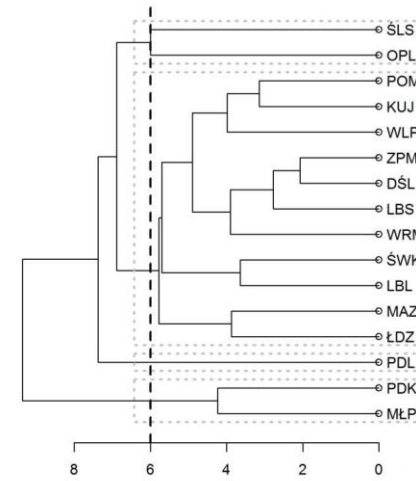
Im. męskie, 1951 r., $h = 6.1$



Im. męskie, 1952 r., $h = 6$

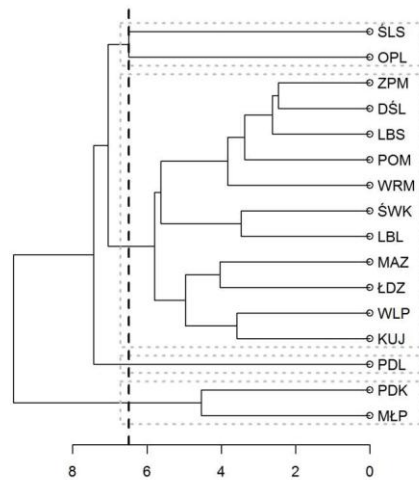


Im. męskie, 1953 r., $h = 5.99$

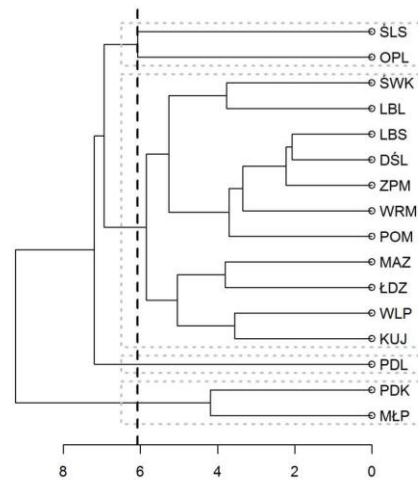


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

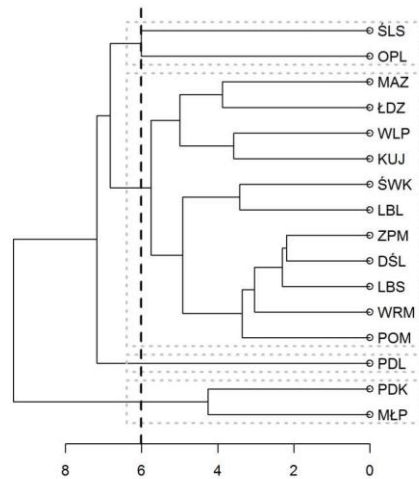
Im. męskie, 1954 r., $h = 6.49$



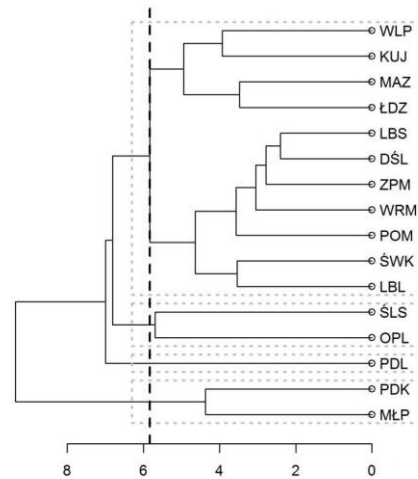
Im. męskie, 1955 r., $h = 6.08$



Im. męskie, 1956 r., $h = 6.02$

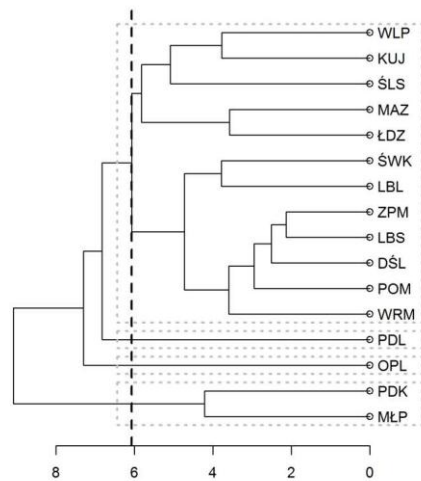


Im. męskie, 1957 r., $h = 5.83$

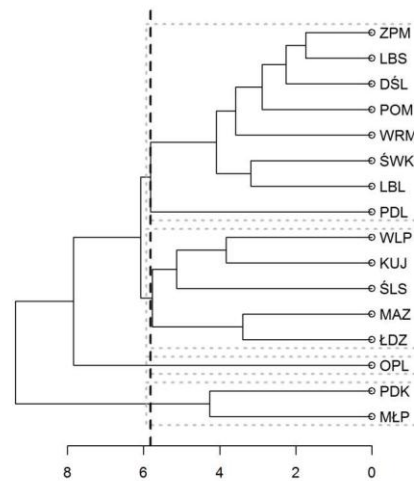


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

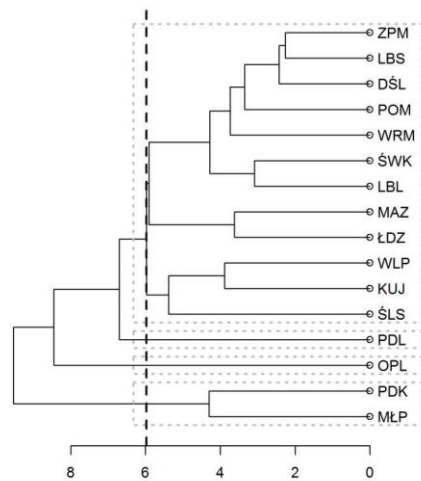
Im. męskie, 1958 r., $h = 6.07$



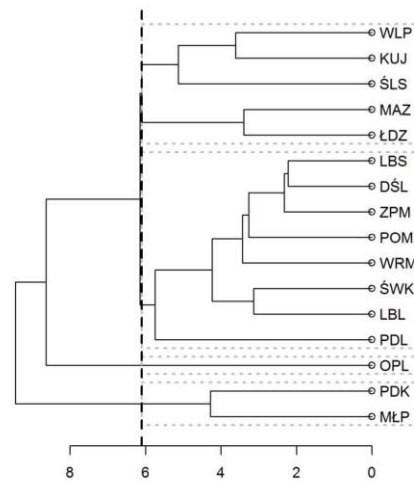
Im. męskie, 1959 r., $h = 5.82$



Im. męskie, 1960 r., $h = 5.97$

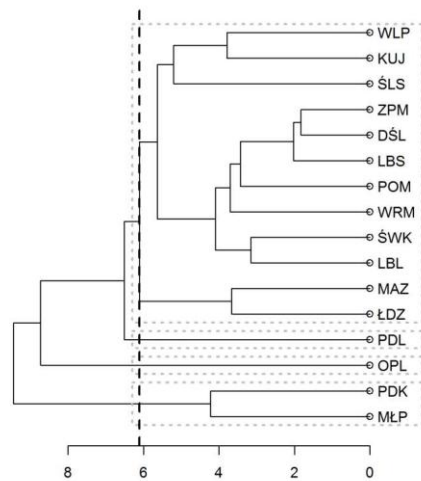


Im. męskie, 1961 r., $h = 6.1$

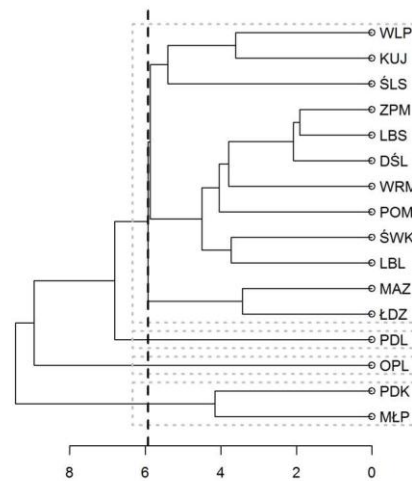


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

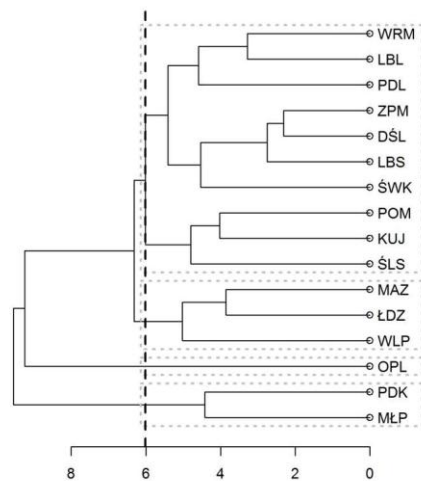
Im. męskie, 1962 r., $h = 6.11$



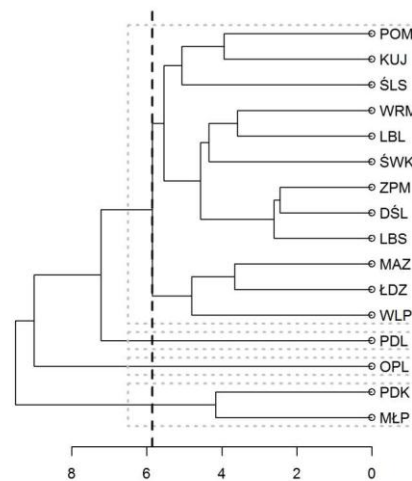
Im. męskie, 1963 r., $h = 5.92$



Im. męskie, 1964 r., $h = 6.02$

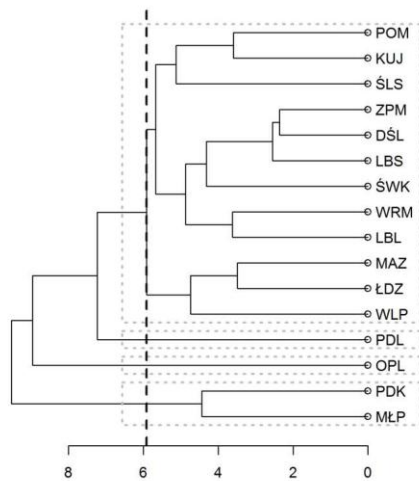


Im. męskie, 1965 r., $h = 5.85$

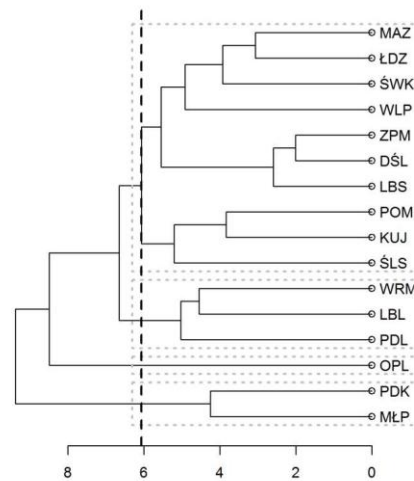


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

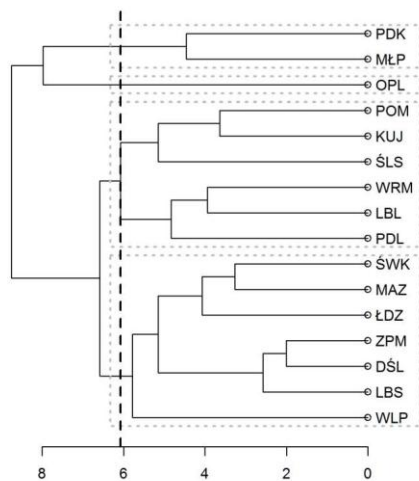
Im. męskie, 1966 r., h = 5.91



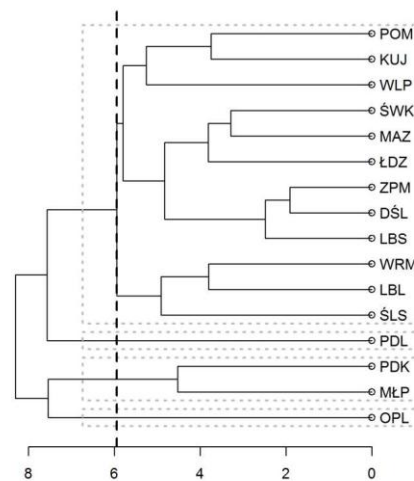
Im. męskie, 1967 r., h = 6.06



Im. męskie, 1968 r., h = 6.07

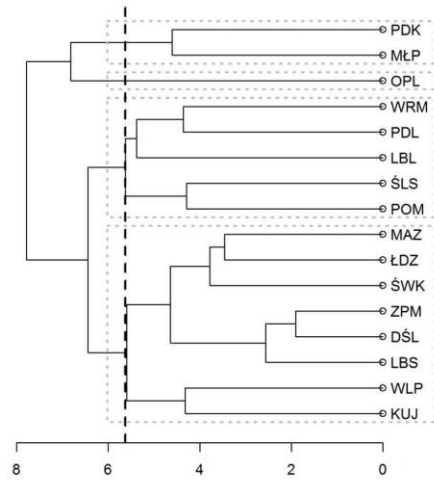


Im. męskie, 1969 r., h = 5.94

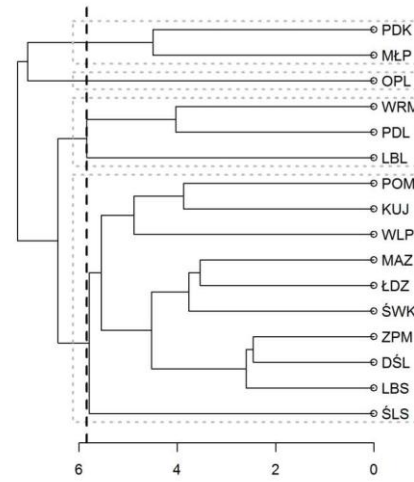


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

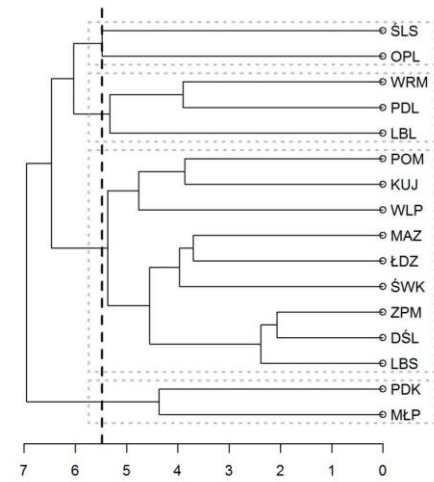
Im. męskie, 1970 r., $h = 5.62$



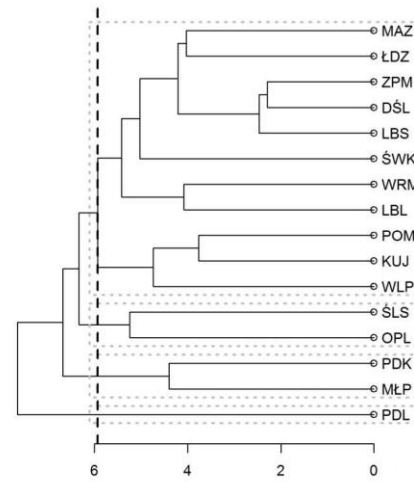
Im. męskie, 1971 r., $h = 5.84$



Im. męskie, 1972 r., $h = 5.48$

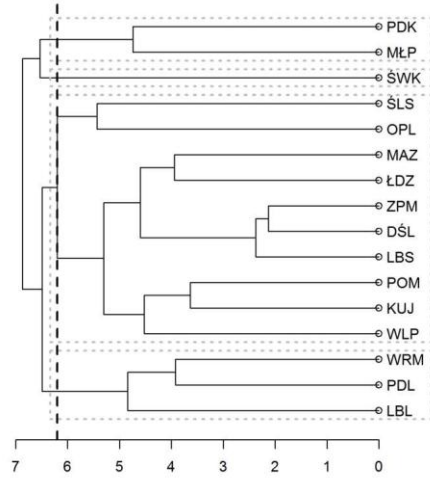


Im. męskie, 1973 r., $h = 5.93$

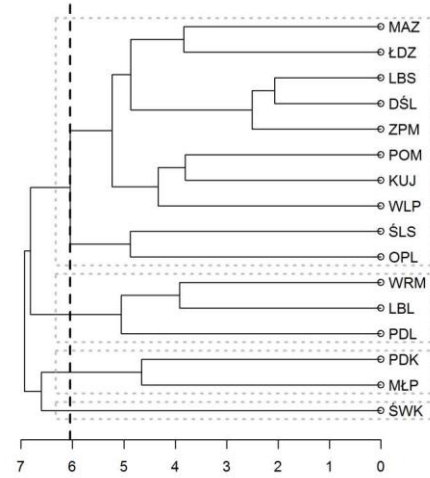


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

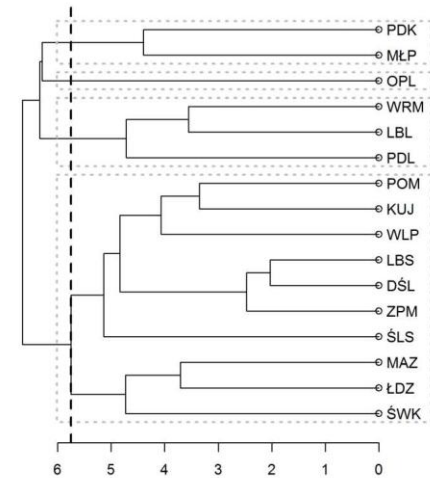
Im. męskie, 1974 r., h = 6.2



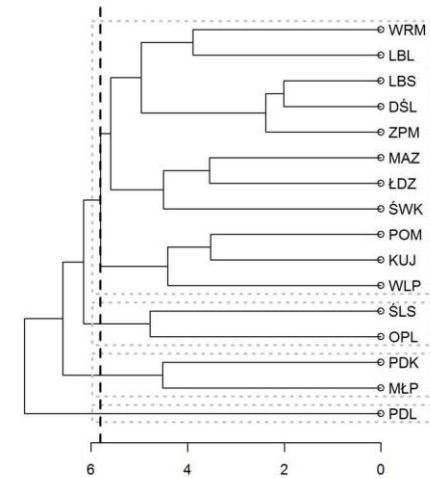
Im. męskie, 1975 r., h = 6.04



Im. męskie, 1976 r., h = 5.75

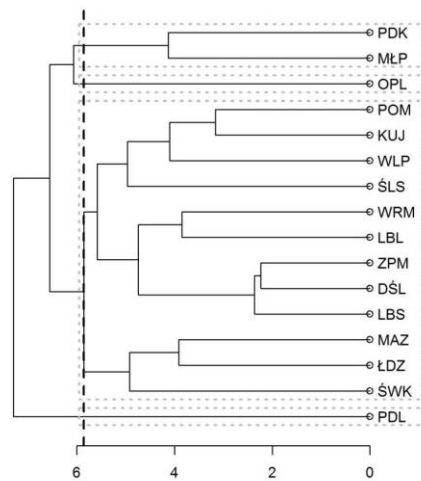


Im. męskie, 1977 r., h = 5.81

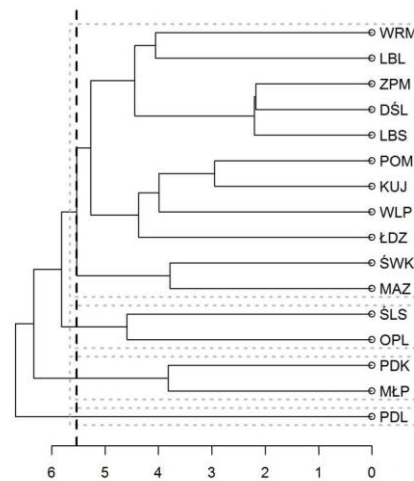


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

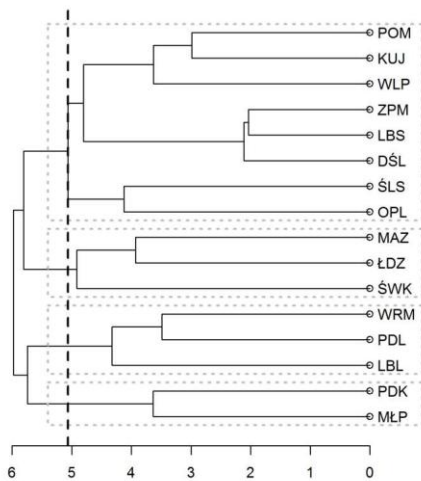
Im. męskie, 1978 r., $h = 5.85$



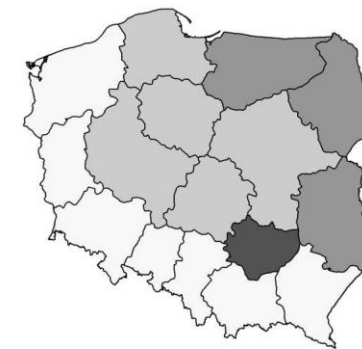
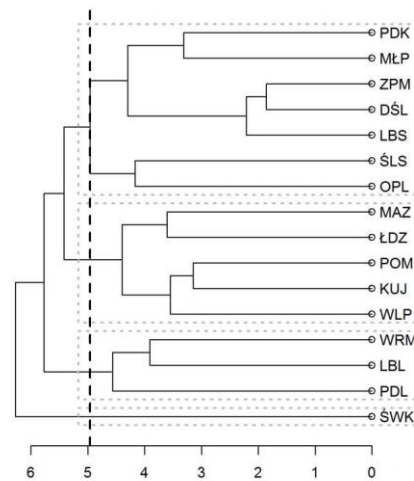
Im. męskie, 1979 r., $h = 5.53$



Im. męskie, 1980 r., $h = 5.06$

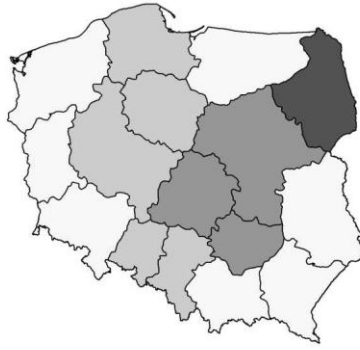
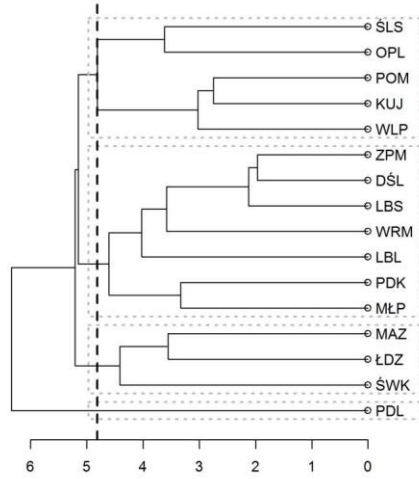


Im. męskie, 1981 r., $h = 4.96$

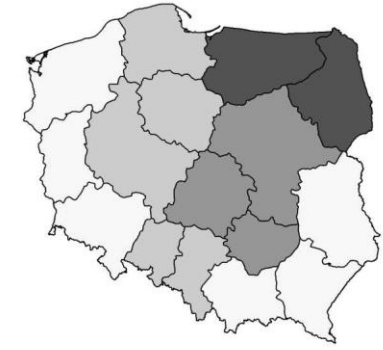
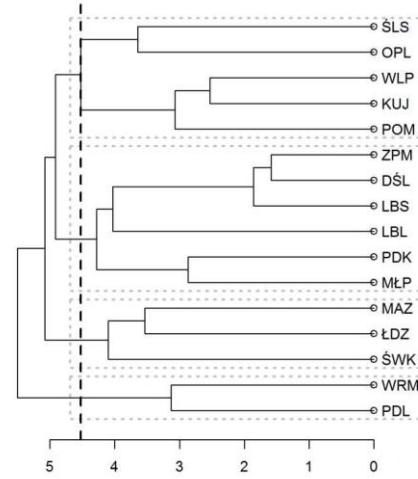


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

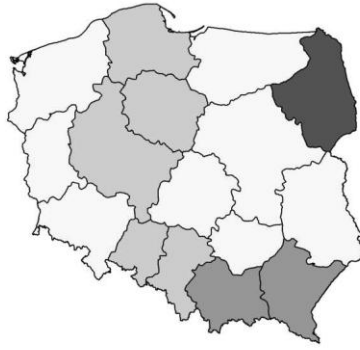
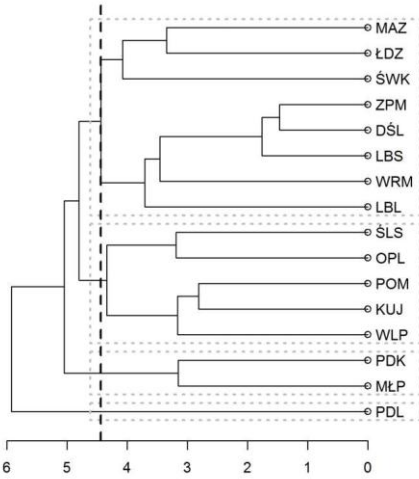
Im. męskie, 1982 r., $h = 4.81$



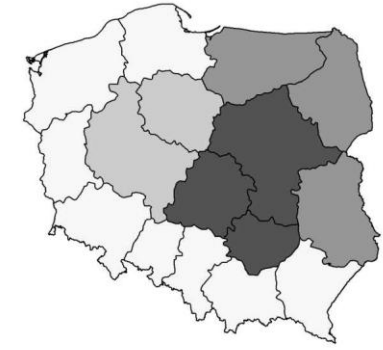
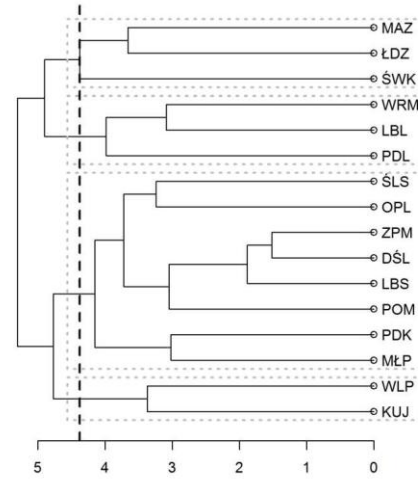
Im. męskie, 1983 r., $h = 4.52$



Im. męskie, 1984 r., $h = 4.44$

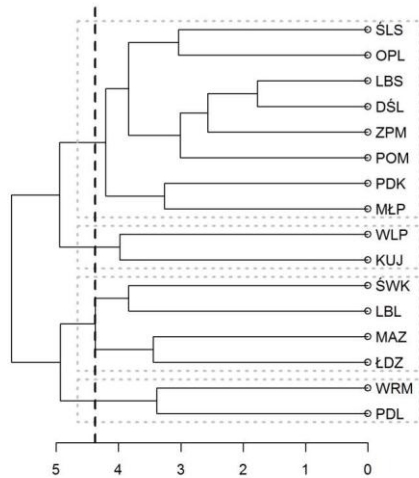


Im. męskie, 1985 r., $h = 4.38$

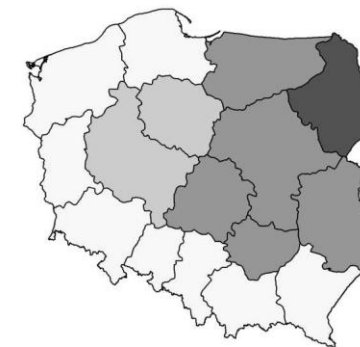
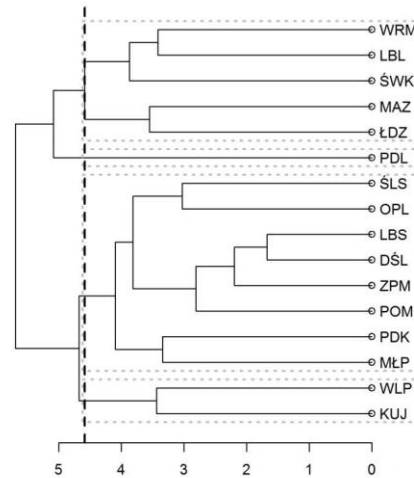


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

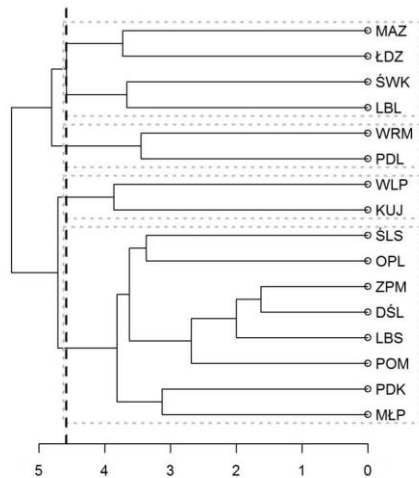
Im. męskie, 1986 r., $h = 4.38$



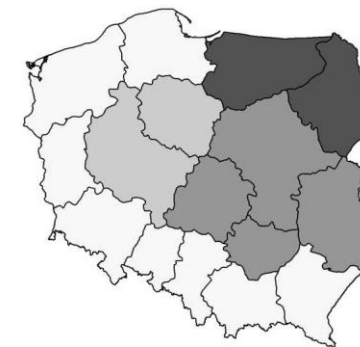
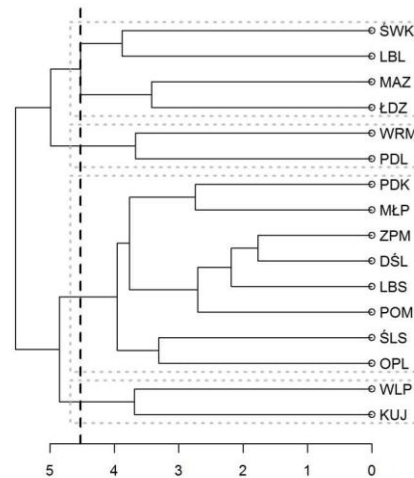
Im. męskie, 1987 r., $h = 4.59$



Im. męskie, 1988 r., $h = 4.59$

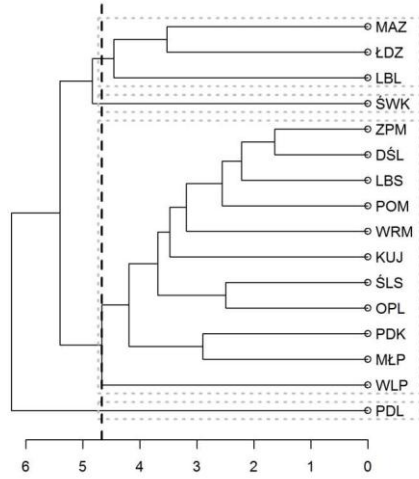


Im. męskie, 1989 r., $h = 4.53$

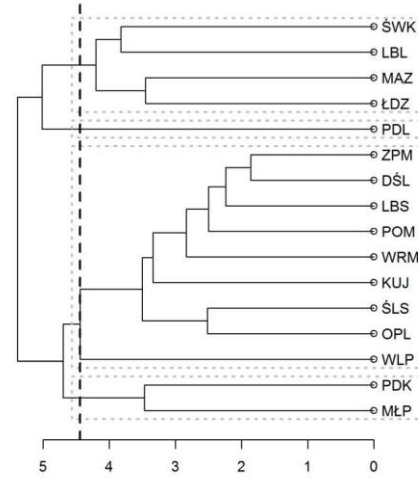


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

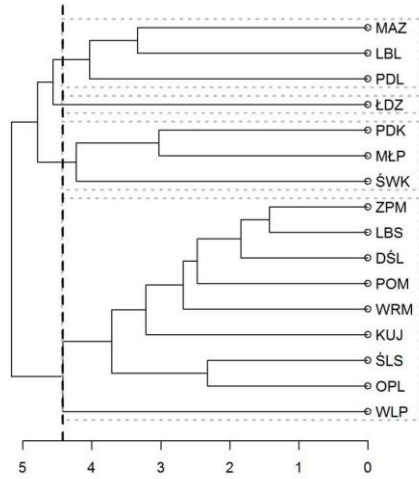
Im. męskie, 1990 r., $h = 4.67$



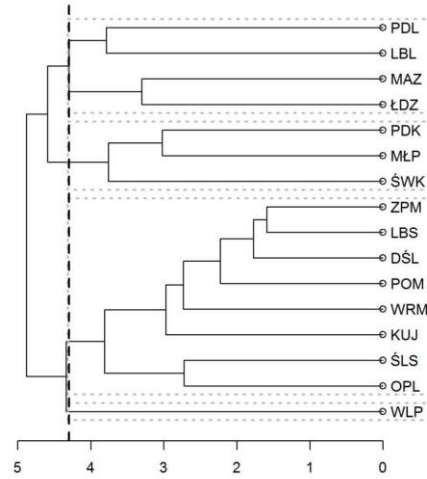
Im. męskie, 1991 r., $h = 4.44$



Im. męskie, 1992 r., $h = 4.42$

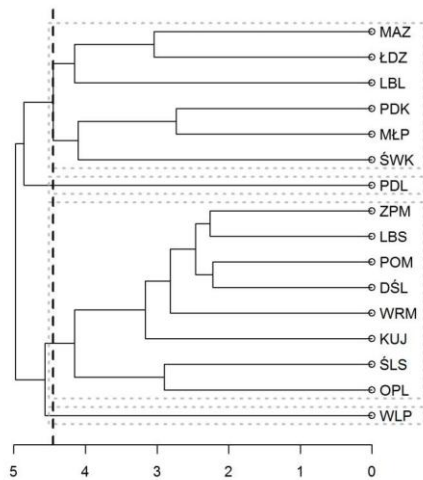


Im. męskie, 1993 r., $h = 4.3$

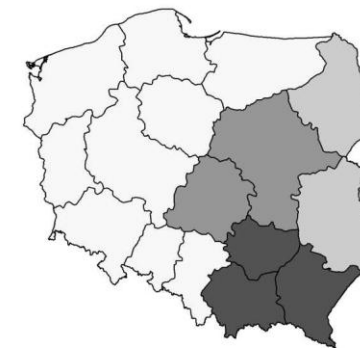
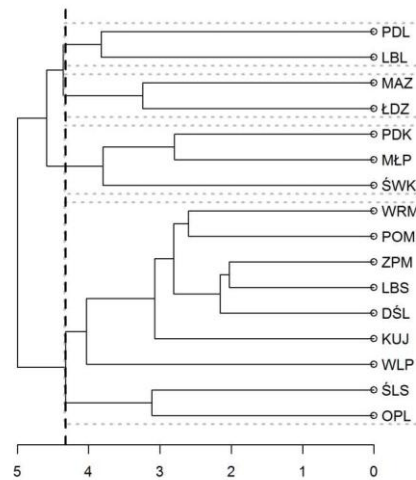


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

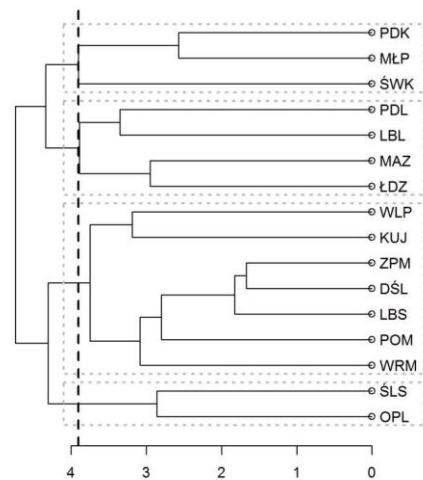
Im. męskie, 1994 r., $h = 4.45$



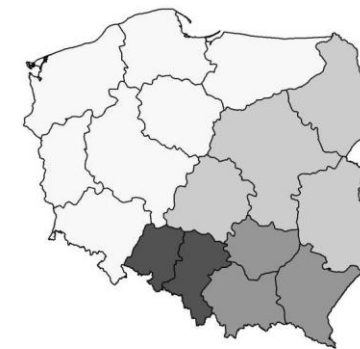
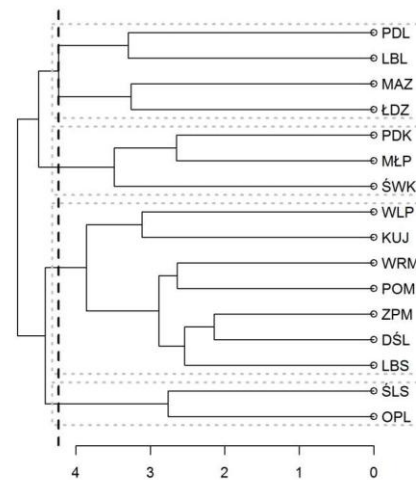
Im. męskie, 1995 r., $h = 4.32$



Im. męskie, 1996 r., $h = 3.9$

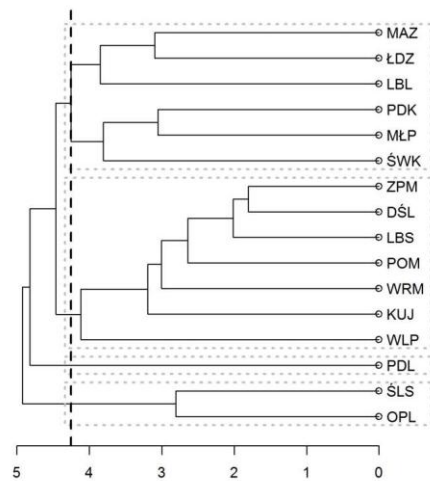


Im. męskie, 1997 r., $h = 4.23$

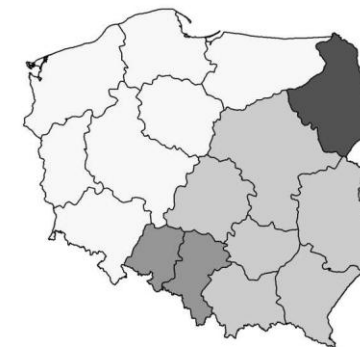
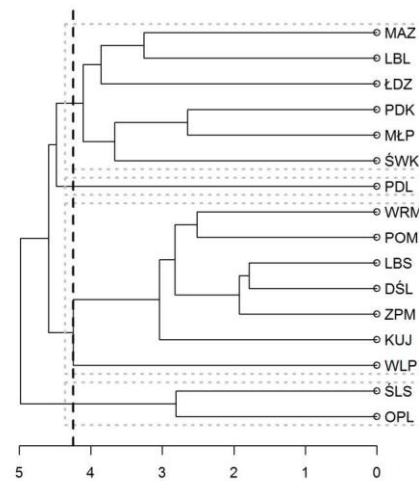


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

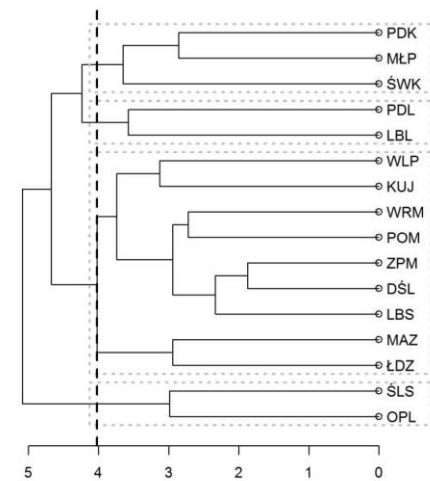
Im. męskie, 1998 r., $h = 4.25$



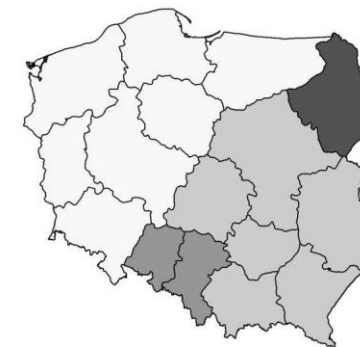
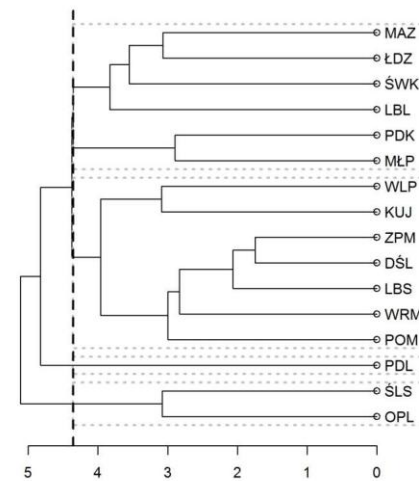
Im. męskie, 1999 r., $h = 4.25$



Im. męskie, 2000 r., $h = 4.02$

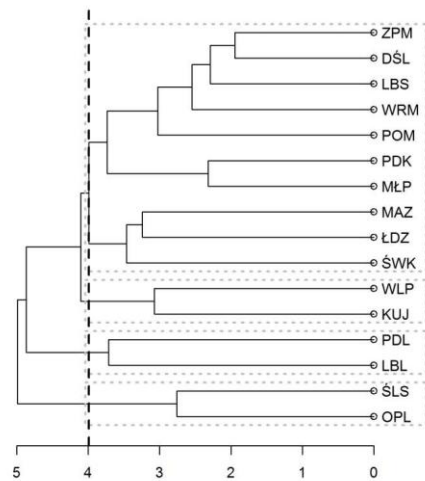


Im. męskie, 2001 r., $h = 4.35$

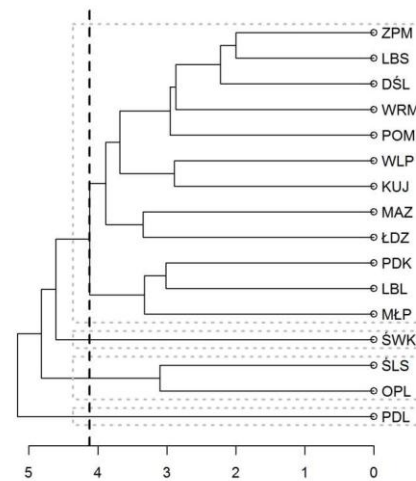


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

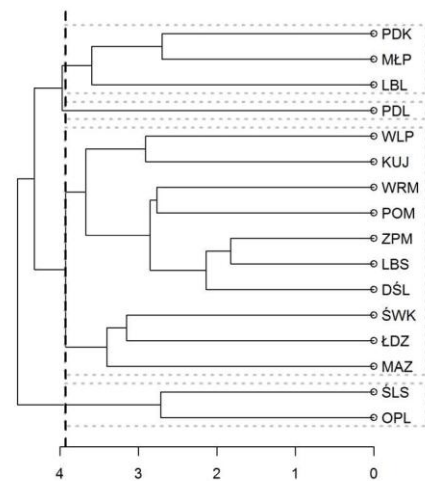
Im. męskie, 2002 r., h = 3.99



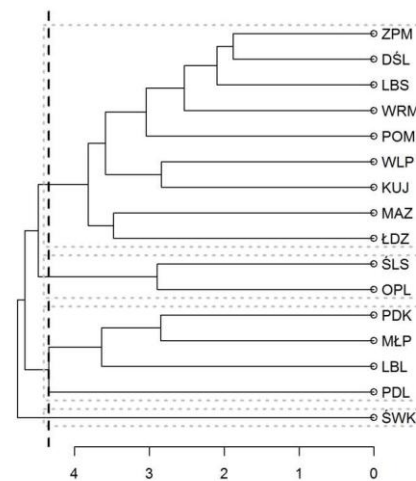
Im. męskie, 2003 r., h = 4.12



Im. męskie, 2004 r., h = 3.93

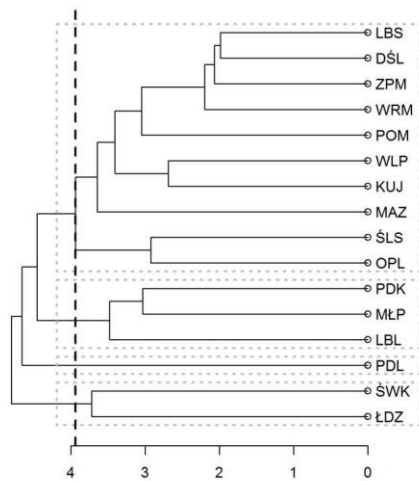


Im. męskie, 2005 r., h = 4.34

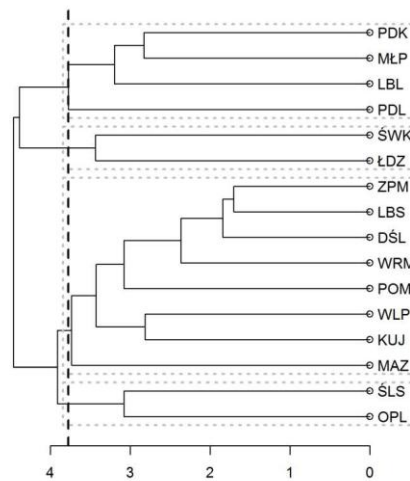


Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

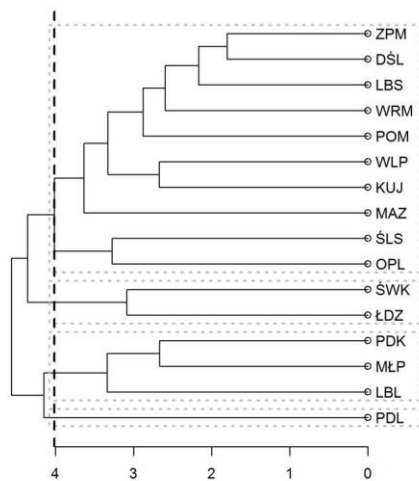
Im. męskie, 2006 r., h = 3.94



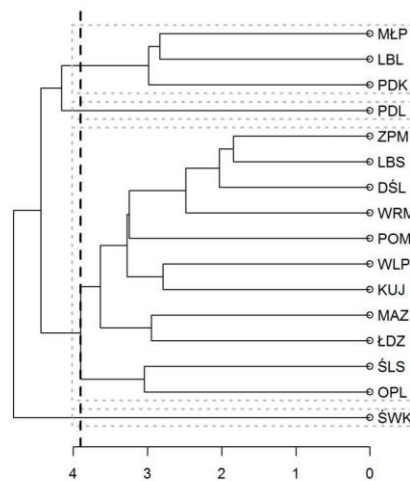
Im. męskie, 2007 r., h = 3.78



Im. męskie, 2008 r., h = 4.01

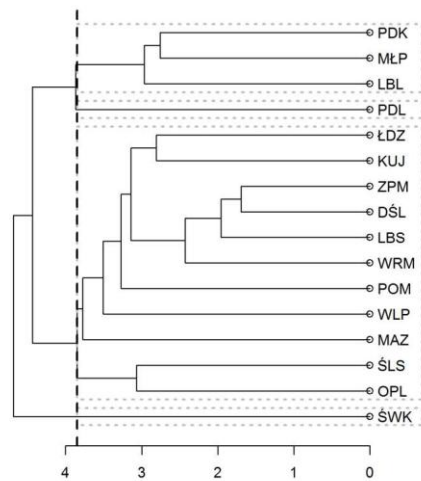


Im. męskie, 2009 r., h = 3.9



Rys. 86 Hierarchiczna analiza skupień – 100 najczęstszych imion męskich. Rozmieszczenie geograficzne czterech największych skupień

Im. męskie, 2010 r., $h = 3.85$



V.6. Wyniki hierarchicznej analizy skupień - imiennictwo na obszarze współczesnych powiatów woj. śląskiego (lata 1931-2010)

Wyniki analizy skupień przeprowadzonej w poprzedniej części ujawniły, że pod kątem układu imion na listach frekwencyjnych województwa często tworzą klastry, które w dużym stopniu odzwierciedlają regiony historyczno-kulturowe, wyznaczone przez granice dawnych państw zaborczych (Austro-Węgier, Prus oraz Cesarstwa Rosyjskiego). Jak już jednak wspomniano, granice współczesnych województw tylko w części pokrywają się z tymi obszarami. W celu weryfikacji, czy w imiennictwie rzeczywiście odzwierciedlają się te dawne podziały, należy zejść do niższego poziomu powiatów lub gmin, które można z dużą dokładnością przyporządkować do któregoś z regionów historycznych. Zbadanie pod tym kątem całego obszaru Polski byłoby zadaniem bardzo pracochłonnym, gdyż każdej gminie i powiatowi konieczne byłoby przypisanie atrybutu przynależności do jakiegoś zdefiniowanego regionu innego niż współczesne województwo, co stanowi raczej zadanie dla historyków i specjalistów od systemów informacji przestrzennej. Postanowiłem w taki szczegółowy sposób zbadać jedynie zróżnicowanie imiennictwa w województwie śląskim, które jest dobrym przykładem obszaru mieszanego – w jego obrębie znajdują się bowiem tereny należące dawniej do wszystkich trzech państw zaborczych.⁸⁶

Do Królestwa Kongresowego będącego częścią Cesarstwa Rosyjskiego należały następujące współczesne powiaty: będziński, częstochowski, kłobucki, myszkowski, zawierciański oraz miasta: Częstochowa, Dąbrowa Górnicza i Sosnowiec. Do Cesarstwa Austro-Węgier należały powiaty bielski, cieszyński i żywiecki oraz miasta Bielsko-Biała i Jaworzno.⁸⁷ W obrębie dawnego państwa pruskiego znajdowały się powiaty: bieruńsko-łędziński, gliwicki, lubliniecki, mikołowski, pszczyński, raciborski, rybnicki, tarnogórski, wodzisławski oraz miasta: Bytom, Chorzów, Gliwice, Jastrzębie-Zdrój, Katowice, Mysłowice, Piekary Śląskie, Ruda Śląska, Rybnik, Siemianowice Śląskie, Świętochłowice, Tychy, Zabrze i Żory. Na rys. 87 widnieje poglądowa mapa, na której zaznaczono przynależność poszczególnych powiatów do jednego z trzech dawnych mocarstw.

⁸⁶ Celowo piszę o przynależności do jednego z tych państw, a nie do zaborów, ponieważ w przypadku samego Górnego Śląska, którego obszar w części pokrywa się ze współczesnym województwem śląskim, nie możemy w żadnym wypadku mówić w kontekście zaborów, gdyż w momencie, kiedy się dokonywały, jego obszar od dawna znajdował się pod panowaniem Prus lub Monarchii Habsburgów (Śląsk Cieszyński).

⁸⁷ Należy przy tym pamiętać, że dawne miasto Bielsko, zachodnia część powiatu bielskiego oraz powiat cieszyński to historyczne tereny Śląska Cieszyńskiego, zaś pozostałe powiaty przynależały do Galicji, która w tym przypadku pokrywa się z historyczną Małopolską.

Rys. 87. Przynależność współczesnych powiatów woj. śląskiego do Cesarstwa Rosyjskiego, Austro-Węgier i Prus przed 1918 r.



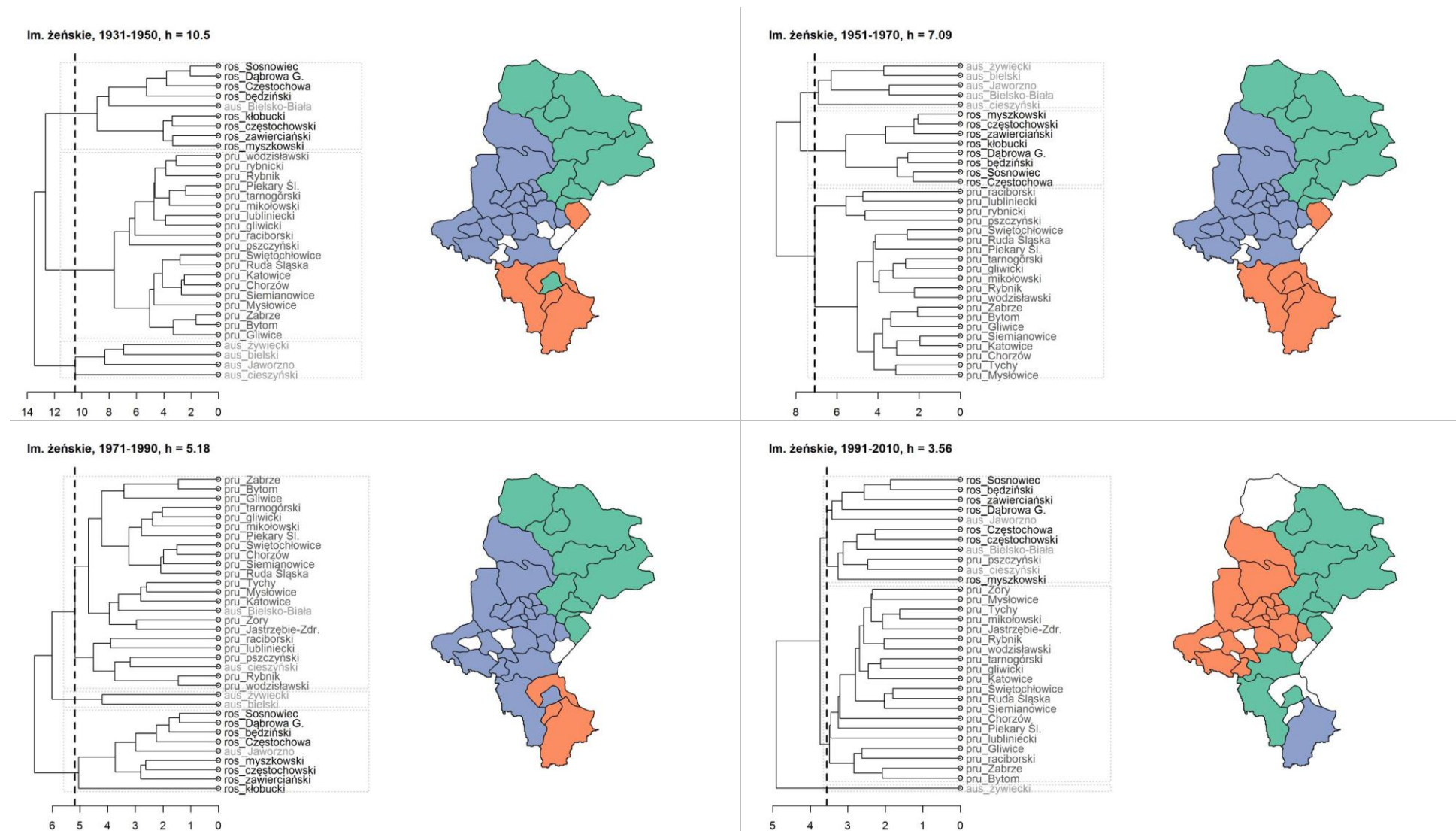
Analizę skupień przeprowadzono przy użyciu tych samych metryk odległości i metod wiązania, jakie zastosowałem w przypadku badania na poziomie województw, jednak przedziały czasowe oraz dobór liczby cech diagnostycznych musiał w tym przypadku ulec modyfikacji. Korpusy imiennicze w pojedynczych powiatach są o wiele mniejsze niż w województwach, dlatego niemożliwe było przeprowadzenie analizy w odniesieniu do pojedynczych lat. Badaniu poddano jedynie 50 najczęstszych imion nadawanych w kolejnych 20-letnich przedziałach czasowych z okresu 1931–2010. Przesunięcie początkowej granicy okresu do 1931 r. było możliwe, gdyż w przypadku większości porównywanych powiatów woj. śląskiego dysponujemy informacjami o imionach nadawanych także przed 1945 r.

Wyniki analizy przedstawione zostały na rys. 88 i 89, które składają się z czterech paneli odpowiadającym przedziałom czasowym 1931–1950, 1951–1970, 1971–1990, 1991–2010, zawierających dwa elementy – dendrogram i mapę. „Liście” dendrogramu opatrzone zostały etykietą z nazwą powiatu, poprzedzoną jednym z trzech prefiksów („aus”, „pru”, „ros”) określającą przynależność do jednego z trzech dawnych mocarstw. Ułatwi to ocenę tego, czy powiaty o podobnym profilu historycznym wykazują podobieństwo również pod względem imiennictwa. Szarą przerywaną linią zaznaczono trzy największe skupienia, zaś czarna

przerwana linia określa wysokość odcięcia niezbędną do ich wydzielenia. Na mapie jednym z trzech kolorów oznaczono przynależność danego powiatu do jednego z trzech największych uzyskanych klastrow (brak koloru oznacza nieuwzględnienie danego powiatu w analizie z powodu zbyt małego korpusu imion).

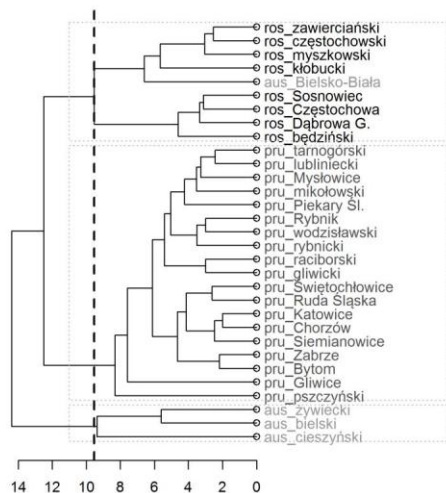
Analizując wyniki klasteryzacji zaprezentowane na rys. 88 i 89, z łatwością możemy zauważyć, że w przypadku dwóch paneli dla okresu 1931–1950 i 1951–1970 granice trzech największych skupień niemal w pełni pokrywają się z granicami historycznych regionów. Jedyne wyjątki stanowią tu dwa miasta należące dawniej do Austro-Węgier, które znalazły się w jednym skupieniu z powiatami zaboru rosyjskiego: Bielsko-Biała (w imiennictwie żeńskim tylko w drugim okresie) i Jaworzno (tylko w imiennictwie męskim), które zresztą z nimi sąsiaduje. W kolejnych dwóch okresach dawne granice nadal są widoczne, choć ulegają nieznacznemu zatarciu. Dotyczy to jednak przede wszystkim imiennictwa żeńskiego, co może być kolejnym dowodem na to, że przy wyborze imion dla dziewcząt rodzice są mniej konserwatywni niż ma to miejsce w przypadku chłopców. O topniejących różnicach w praktykach imienniczych między wszystkimi powiatami świadczy z pewnością zmniejszająca się regularnie z okresu na okres wysokość odcięcia trzech głównych skupień (wartość h). W przypadku imion żeńskich w pierwszym okresie było to 10,5, zaś w ostatnim tylko 3,56, zaś w imiennictwie męskim odpowiednio 9,53 i 3,71. Obserwacje te pokrywają się z wnioskami E. Rudnickiej-Firy (2011: 205) o stopniowym zacieraniu się różnic między imiennictwem Górnego Śląska i Zagłębia Dąbrowskiego. Nie zmienia to jednak faktu, że w imiennictwie powiatów woj. śląskiego widoczne są nadal różnice odzwierciedlające przynależność do różnych regionów historyczno-kulturowych. Pośrednio daje to też asumpt do przypuszczenia, że geograficzne zróżnicowanie zaobserwowane w toku analizy skupień na poziomie województw odzwierciedla te same podziały.

Rys. 88. Hierarchiczna analiza skupień – 50 najczęstszych imion żeńskich w powiatach województwa śląskiego. Rozmieszczenie trzech największych skupień

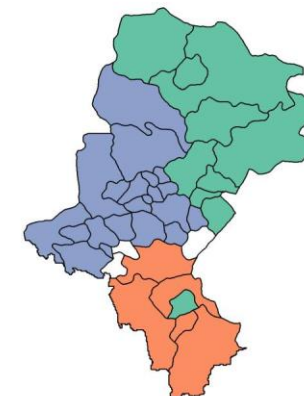
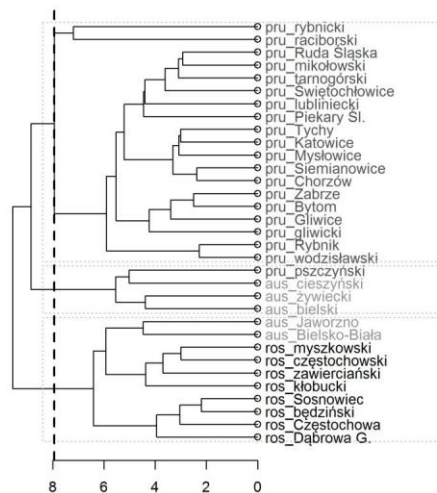


Rys. 89. Hierarchiczna analiza skupień – 50 najczęstszych imion męskich w powiatach województwa śląskiego. Rozmieszczenie trzech największych skupień

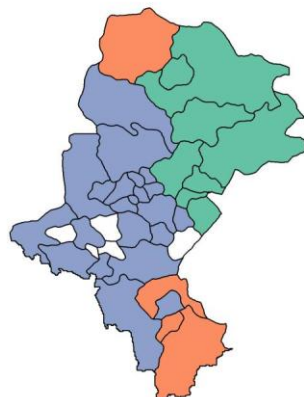
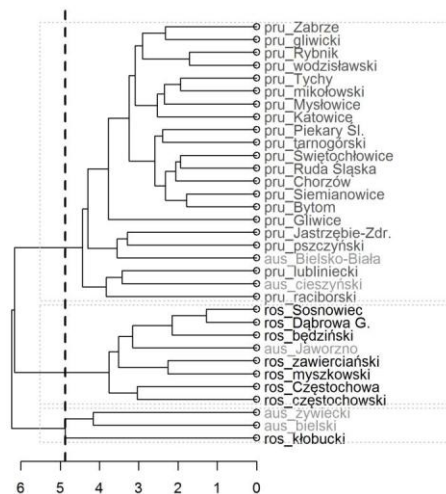
Im. męskie, 1931-1950, $h = 9.53$



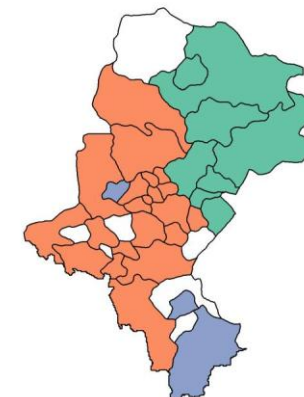
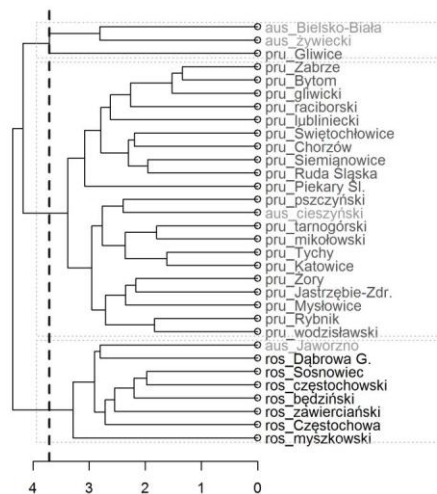
Im. męskie, 1951-1970, $h = 7.94$



Im. męskie, 1971-1990, $h = 4.87$



Im. męskie, 1991-2010, $h = 3.71$



VI. Zakończenie

Badania przeprowadzone przeze mnie w rozdziałach III, IV i V przyniosły odpowiedzi na pytania postawione we wstępie pracy, jednak od razu należy zaznaczyć, że nie zawsze są to odpowiedzi jednoznaczne. Pytania te dotyczyły tego, czy w zakresie wybranych cech ilościowych występowało czasowe i przestrzenne zróżnicowanie systemu imienniczego warunkowane takimi czynnikami, jak przynależność czasowa, przynależność terytorialna czy płeć nazywanych osób. Zaobserwowane zostały pewne charakterystyczne zjawiska zarówno w zakresie czasowego, jak i przestrzennego zróżnicowania imiennictwa, a także istotne różnice między imiennictwem męskim i żeńskim, jednak w zależności od typu przeprowadzonej analizy charakter tych zjawisk może być odmienny.

W toku analizy dotyczącej realizacji wybranych parametrów rozkładu częstości wykazano, że w imiennictwie ogólnopolskim w zakresie każdego z badanych wskaźników na przestrzeni całego rozpatrywanego okresu mogliśmy zaobserwować dużą zmienność ich wartości, której kształt mógł być warunkowany czynnikami pozajęzykowymi. Wiele badanych wskaźników przejawiało bowiem bardzo podobne tendencje, których punkty zwrotne zbiegały się w czasie z istotnymi w historii Polski zdarzeniami. Niezaprzeczalny wpływ tych wydarzeń na przemiany społeczne, znalazł odbicie także w zmianach w zachowaniach nazewniczych. Mam tu na myśli wyraźne różnice w strukturze kwantytatywnej imiennictwa w okresie sprzed 1945 r., w czasie PRL oraz w okresie po transformacji ustrojowej 1989 r. Prawie zawsze w okresie międzywojennym i aż do połowy lat czterdziestych, a więc do końca II wojny światowej, mieliśmy do czynienia ze wzrostem bogactwa zasobów imienniczych w kolejnych latach. Po 1945 r. niemal zawsze następował jego znaczny spadek i aż do początku lat 80., zatem niemal przez cały okres PRL, utrzymywał się na stosunkowo niskim poziomie. Można to odczytywać jako przejaw unifikacji imiennictwa w okresie PRL, co odnotowała w badaniach nad imiennictwem Białegostoku Z. Abramowicz (1993: 370). Zjawisko to wpisuje się w szersze procesy unifikacji społecznej i kulturowej w tym czasie, może być też przejawem konserwatyzmu obyczajowego PRL.

Od lat 80. ponownie mamy do czynienia ze wzrostem różnorodności imiennictwa, który uległ znacznemu i trwałemu przyspieszeniu na przełomie lat 80. i 90. Nietrudno zauważyć, że początek tego procesu przypada na czas kształtowania się nowej świadomości obywatelskiej związanej z ruchem Solidarności (Dubisz 1995: 80). Moment jego intensyfikacji pokrywa się z czasem transformacji systemu gospodarczo-politycznego 1989 r., której konsekwencją był

proces daleko idącego rozwarstwienia społeczeństwa polskiego, różnicujący jego style życia (Dyczewski 1995: 152). Otwarcie się na procesy globalizacji, przemiany ekonomiczne, obyczajowe, pluralizacja życia społecznego, wzmożone migracje czy w końcu przeobrażenia i rozwój środków masowego przekazu – wszystkie te czynniki mogły mieć wpływ na odnotowaną w toku analizy zmianę ilościowego profilu polskiego systemu imienniczego po roku 1989.

Najwyraźniej wspomniane tendencje można było zaobserwować w przypadku wielkości słownika, udziału najrzadszej leksyki, wskaźnika stereotypowości i koncentracji Mistrika oraz parametru α . Warto jednak pamiętać, że niektóre wskaźniki wykazywały nieco odmienne tendencje polegające albo – jak w przypadku wskaźnika koncentracji Lorenza w imiennictwie męskim – tendencję diametralnie inną. Trzeba zresztą stanowczo podkreślić, że w żadnym wypadku nie występuje pełna zgodność między poszczególnymi wskaźnikami – choć między niektórymi z nich zachodzi silna korelacja, to nigdy nie jest to korelacja zupełna. Należy mieć to na uwadze przy wszelkich próbach opisu zbiorów imiennicznych (ale także i tekstów języka naturalnego) przy ich pomocy.

Analiza parametrów rozkładu częstości w imiennictwie na obszarach współczesnych województw dowiodła, że istnieje nie tylko czasowe zróżnicowanie imiennictwa pod kątem struktury ilościowej, ale także przestrzenne. Co więcej, okazało się, że w przypadku większości badanych wskaźników układ geograficzny otrzymanych w toku analizy wartości wskaźników odzwierciedla w pewnym stopniu dwa najbardziej trwałe typy podziału regionalnego Polski obserwowane w wielu dziedzinach życia społecznego. Chodzi mianowicie o charakterystyczne „pęknięcie” na wschód i zachód, a także nieco bardziej szczegółowy podział, uwzględniający trwałość granic dawnych państw zaborczych oraz terytoriów uzyskanych po 1945 r. Bodaj najbardziej wyrazistym zjawiskiem, zwłaszcza w początkowym okresie (1951–60) było to, że najmniejszą różnorodnością imiennictwa odznaczały się województwa małopolskie i podkarpackie, zaś na przeciwległym biegunie pod tym względem plasowało się imiennictwo województw skrajnie zachodniej części Polski. Pokrywa się to z obserwacjami socjologów na temat szczególnego nasilenia postaw konserwatywnych wśród mieszkańców dawnej Galicji i podatności na procesy modernizacji mieszkańców Ziemi Zachodnich. Taki stan ma wiele przyczyn – w odróżnieniu od „osiadłej” Galicji Ziemia Zachodnie zamieszkuje populacja ukształtowana w wyniku migracji, wymieszana genetycznie i kulturowo, nie bez znaczenia jest fakt, iż ludność tego drugiego

obszaru jest znacznie młodsza niż pozostałych regionów (Jałowiecki 1996: 6–7, Bartkowski 2003: 423, Halamska 2007: 235).

O funkcjonowaniu wspomnianych typów regionalizacji w imiennictwie świadczą również wyniki uzyskane przy pomocy hierarchicznej analizy skupień, której celem było zbadanie podobieństwa obszarów współczesnych województw pod kątem układu najczęstszych imion na listach frekwencyjnych. Najbardziej wyraziste skupienia, które zostały wydzielone w toku analizy, pokrywały się w dużej mierze z regionami o różnych profilach społeczno-histerycznych, takimi jak Galicja, Królestwo Kongresowe, Ziemie Zachodnie czy należące w przeszłości do Prus Górny Śląsk, Wielkopolska i Pomorze. Pokazuje to, że procesy „długiego trwania”, kształtujące do dziś zróżnicowanie regionalne Polski, w którym uwidacznia się wpływ wielkich wydarzeń historycznych (Bartkowski 2003: 421–425), odbijają się również w zróżnicowaniu ilościowym imiennictwa. Jeszcze dobitniej dowiodło tego badanie przeprowadzone na materiale imienniczym powiatów woj. śląskiego, gdzie granice największych trzech klastrów niemal w pełni pokrywały się z granicami dawnych państw zaborczych.

Analiza ujawniła też, że pod kątem częstości dobieranych przez rodziców imion na tle pozostałej części kraju często wyraźnie odbijają się Podlasie i Opolszczyzna, czyli obszary o dużym udziale mniejszości narodowych.

Należy jednak podkreślić, że zaobserwowane w badaniach parametrów rozkładu częstości oraz układu imion na listach frekwencyjnych podziały regionalne, warunkowane różnicami w strukturze kwantytatywnej imiennictwa, z upływem czasu tracą na wyrazistości i ulegają stopniowej niwelacji. Dystans, jaki dzielił imiennictwo poszczególnych obszarów na przestrzeni rozpatrywanego okresu, regularnie malał, a wyraziste początkowo zwarte obszary ulegają przeobrażeniom. Widoczne są tu analogie ze zjawiskiem integracji językowej obserwowanej w polszczyźnie, przejawiającej się w zacieraniu tradycyjnych podziałów regionalno-kulturowych i dawnych regionalizmów oraz w bardzo intensywnym zanikaniu od lat 60. dialektów ludowych w wyniku migracji ludności ze wsi do miast, oddziaływania systemu oświatowego czy normatywizacji wzorców komunikatywnych (Dubisz 1995: 81).

Ciekawe, choć poniekąd oczekiwane wyniki przyniosło badanie pod kątem różnic w strukturze ilościowej imiennictwa ośrodków wielkomiejskich na tle obszarów o niższym poziomie urbanizacji. W zdecydowanej większości przypadków duże miasta odznaczały się znacznie bardziej różnorodnym ilościowo imiennictwem od obszarów prowincjonalnych. Szczególnie uderzający był poziom różnic między Warszawą a pozostałym terenem

województwa mazowieckiego. Warszawa odróżniała się nie tylko na jego tle, ale też pozostałych badanych miast, jeśli chodzi o poziom różnic między imiennictwem męskim i żeńskim – zwykle był on tam najniższy. Na przestrzeni badanego okresu różnice między porównywanymi obszarami o różnym poziomie urbanizacji ulegały stopniowemu zmniejszaniu (choć znów w najmniejszym stopniu dotyczyło to miasta stołecznego), co koresponduje z obserwacją niwelacji różnic językowych między mieszkańcami miast i prowincji (Warchała 2003: 160–161).

Wyjątkowość imiennictwa Warszawy wynika z demograficznej specyfiki tego miasta. Tak w latach powojennych, jak i dziś jest to najczęstszy cel migracji zarówno wewnętrznych, jak i spoza kraju – to tu osiedla się największa liczba obcokrajowców. W całym XX w. stołeczne miasto wyróżniało się m.in. tym, że na rozwój jego ludności miało wpływ przede wszystkim osiedlanie się w nim przybyszy z różnych regionów kraju – w pierwszych latach powojennych był to czynnik istotniejszy nawet od przyrostu naturalnego (Kosiński 1968: 16–24, Kałuża-Kopias 2014: 52). Sprawia to, że ludność Warszawy jest zdecydowanie najbardziej kulturowo i społecznie zróżnicowana, co musiało znaleźć odbicie w jego imiennictwie.

Powyższe ogólne obserwacje poczynione na podstawie wyników analizy czasowego i przestrzennego zróżnicowania imiennictwa w zakresie parametrów rozkładu częstości imion w zbiorach oraz ich układu na listach frekwencyjnych tylko w niewielkim stopniu pokrywają się z wnioskami płynącymi z badania tempa rotacji w najczęstszym imiennictwie. W przypadku imiennictwa ogólnopolskiego stwierdzono ewidentne przyspieszenie tempa rotacji na przestrzeni całego rozpatrywanego okresu, jednak nie ujawniły się tu konkretne cezury wyznaczające ramy tego procesu. Jego wyrazistość była zależna od dobranej liczby najczęstszych imion oraz szerokości przedziału czasowego, w jakim było mierzone. Zdecydowanie bardziej wyraźny i regularny wzrost dynamiki zmian mogliśmy zaobserwować przy stosowaniu wskaźnika tempa rotacji w dłuższych, 10-letnich przedziałach czasowych. Kolejny raz potwierdza to, że ocena danego zjawiska zależy w dużej mierze od zastosowanej metody. Niemniej w przypadku każdej kombinacji liczby najczęstszych imion i długości przedziału czasowego mogliśmy zaobserwować trend rosnący w tempie rotacji, co pokrywa się z wynikami badań nad imiennictwem innych krajów Europy oraz Ameryki Północnej. Możemy zatem przypuszczać, że również w Polsce w coraz większym stopniu wpływ na zachowania nazewnicze mają mechanizmy mody. Zwyczaj nazewniczy, polegający

na replikacji dotychczasowego stanu, zostaje wypierany przez indywidualne gusta rodziców (Liebersson 2000: 66–68).

W przypadku badania tempa rotacji w odstępach 5-letnich w latach 1950–2010 na obszarze współczesnych województw, choć odnotowaliśmy istnienie samego zróżnicowania wartości pomiędzy poszczególnymi obszarami, to jednak wydaje się mieć ono charakter przypadkowy i w żadnym stopniu nie odzwierciedla wzorców regionalnych, jakie zostały zaobserwowane przy badaniu pozostałych zagadnień. Nie odnotowano też żadnych istotnych różnic pomiędzy ośrodkami wielkomijskimi i sąsiadującymi z nimi obszarami o mniejszym stopniu urbanizacji. Ciekawe obserwacje przyniosło natomiast badanie dynamiki zmian w skali rocznej, które ujawniło bardzo wyraźne zmiany w imiennictwie w trakcie II wojny światowej. Bardziej szczegółowa analiza wykazała, że istotne zmiany w tempie rotacji zaszły tylko na obszarze niektórych niemieckich stref okupacyjnych i miały bezpośredni związek z polityką językową władz okupacyjnych.

Istotnym zagadnieniem, które stale pojawiało się we wszystkich analizach była kwestia różnic ilościowych pomiędzy imiennictwem żeńskim i męskim. Badanie parametrów rozkładu częstości w imiennictwie ogólnopolskim ujawniło, że w przypadku niektórych wskaźników (wielkości słownika, najrzadszej leksyki, wskaźników koncentracji i stereotypowości Mistríka) zasób imion użytych do nominacji osób płci żeńskiej był niemal zawsze bardziej różnorodny, co pokrywa się z wnioskami innych badaczy tak polskich, jak zagranicznych, o bardziej konserwatywnym podejściu rodziców do nadawania imion chłopcom. W zakresie pozostałych parametrów mieliśmy do czynienia z sytuacją naprzemienną. W przypadku większości wskaźników odnotowano też tendencję do zmniejszania się wraz z upływem czasu różnic pomiędzy imiennictwem żeńskim i męskim, która ujawniła się również przy analizie parametrów rozkładu częstości na poziomie województw oraz miast. Natomiast badanie tempa rotacji nie wykazało żadnych istotnych różnic w tym zakresie.

Zagadnienia związane ze strukturą kwantytatywną polskiego imiennictwa i jej przeobrażeniami to obszary rzadko eksplorowane w polskiej onomastyce. Niniejsza praca w pewnym stopniu wypełnia tę lukę. Starłem się w niej pokazać, że badanie systemu imienniczego przez pryzmat jego cech ilościowych może dostarczyć nam cennych informacji o jego czasowej i przestrzennej wariacji w kontekście różnych uwarunkowań historyczno-społecznych. Wykorzystanie wielkoskalowego materiału badawczego i jego analiza przy

użyciu odpowiednich metod kwantytatywnych pozwoliły niejako z oddali spojrzeć na te kwestie i dostrzec zjawiska, które często w szczegółowych badaniach wycinkowych mogą nie być dostatecznie widoczne. Zaprezentowanie możliwości wykorzystania niektórych metod statystycznych z dziedziny statystyki leksykalnej czy stylometrii do analizy porównawczej zbiorów imienniczych było zresztą moim pośrednim celem. Wydaje mi się, że zaproponowane przeze mnie rozwiązania mogą znaleźć zastosowanie w dalszych badaniach. Należy jednak pamiętać, że choć starałem się zdywersyfikować metody analizy i prezentacji jej wyników, to stanowią one jedynie ograniczony wybór i należy poszukiwać innych, być może bardziej efektywnych sposobów opisu właściwości ilościowych imiennictwa.

Oczywiście zdaję sobie sprawę, że zakres zagadnień poruszanych przeze mnie w pracy stanowi jedynie fragment złożonej i obszernej problematyki imiennictwa i jego funkcjonowania na tle społecznym. Z uwagi na niewielką ilość dodatkowych informacji, dotyczących nosicieli imion w materiale badawczym, zagadnienia socjolingwistyczne, obejmujące kwestię czasowego, regionalnego czy też demograficznego zróżnicowania imiennictwa w aspekcie ilościowym zostały potraktowane w sposób ogólniejszy.

W mojej pracy zasadniczo też nie odnosiłem się do cech jakościowych imiennictwa. Nie oznacza to oczywiście, że kwestie związane z formą językową, pochodzeniem czy genezą imion uważam za mniej istotne. Analiza ilościowa i jakościowa, w moim przekonaniu, powinny być względem siebie komplementarne i tylko łącząc te dwie perspektywy, możemy uzyskać w pełni satysfakcjonujące odpowiedzi na pytania dotyczące wariacji polskiego imiennictwa. Naturalną kontynuacją przeprowadzonych przeze mnie badań powinno być zatem przyjrzenie się czasowej i przestrzennej dystrybucji cech jakościowych antroponimów składających się na polski system imienniczy.

Pożądane byłoby również poszerzenie perspektywy czasowej obu typów badań. Charakter materiału badawczego wykorzystanego w niniejszej pracy wymógł jej zawężenie do XX w. i początku XXI w. Pewne szanse na przeprowadzenie wielkoskalowych analiz na materiale imienniczym z wcześniejszych okresów stwarzają coraz intensywniejsze prace w zakresie digitalizacji i udostępniania potężnych zasobów metrykalnych z polskich archiwów. Nic natomiast nie stoi na przeszkodzie i jest wręcz wskazane, by przedstawione w pracy badania systematycznie kontynuować na materiale imienniczym pozyskiwanym na bieżąco z bazy PESEL w kolejnych latach. W związku z coraz większą dostępnością baz danych rejestrujących imiennictwo innych krajów zasadne wydaje się też wysunięcie postulatu

przeprowadzenia spójnych analiz porównawczych na poziomie międzynarodowym, w których można by wykorzystać zaproponowane przeze mnie metody.

VII. Bibliografia

- Abramowicz Z. (1993). Imiona chrzestne białostoczan w aspekcie socjolingwistycznym (lata 1885–1985). Białystok.
- Abramowicz Z. (1994). Lingwistyczna i ekstralingwistyczna motywacja wyboru imienia w Białymstoku w latach 1885–1985, [w:] Jazyková a mimojazyková stránka vlastných mien, red. E. Krošláková, Bratislava – Nitra, s. 152–158.
- Adamiak K. (2013). Imiona chrzestne dzieci chłopskich okolic Fajslawic w drugiej połowie XVIII w, „Roczniki Historii Kościoła”, t. 5(60), s. 139–179.
- Alhaug G. (2002). Personal names in North Norway, „Onoma” 37, s. 301–322.
- Allerhand M. (1899). Prawo imion. Lwów.
- Anderson T. (1985). Unique and Common First Names of Males and Females, „Psychological Report” 57, s. 204–207.
- Baayen R. H. (2001). Word frequency distributions. Dordrecht.
- Baayen R. H. (2008). Analyzing Linguistic Data. A Practical Introduction to Statistics using R. New York.
- Baayen R. H. (2013). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 1.4.1. <http://CRAN.R-project.org/package=languageR>
- Bajerowa I. (red.) (1996). Język polski czasu drugiej wojny światowej (1939–1945). Warszawa.
- Balicki A. (2009). Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne. Gdańsk.
- Banderowicz K. (2015). Imiona poznaniaków u progu XVII stulecia (na materiale Akt sądu wójtowskiego). Poznań.
- Barnard A., Spencer J. (red.) (2011). Encyklopedia antropologii społeczno-kulturowej. Warszawa.
- Barry III H., Harper A. S. (2010). Racial and Gender Differences in Diversity of First Names, „Names” 58, s. 47–54.

- Bartkowski J. (2003). *Tradycja i polityka. Wpływ tradycji kulturowych polskich regionów na współczesne zachowania społeczne i polityczne*. Warszawa.
- Bartol-Jarosińska (1991). Społeczne osie wariantywności, „Socjolingwistyka” 1991, nr 11, s. 57–65.
- Barucca P., Rocchi J., Marinari E., Parisi G., Ricci-Tersenghi F. (2015). Cross-correlations of American baby names, „Proceedings of the National Academy of Sciences of the United States of America” 112(26), s. 7943–7947.
- Berger J., Bradlow E. T., Braunstein A., Zhang Y. (2012). From Karen to Katie: Using Baby Names to Understand Cultural Evolution, „Psychological Science” XX(X), s. 1–7.
- Biecek P., Trajkowski K. (2011). Na przełaj przez Data Mining z pakietem R [on-line: <http://www.biecek.pl/NaPrzelajPrzezDataMining>].
- Bivand R., Lewin-Koh N. (2016). *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.8-39, <http://CRAN.R-project.org/package=maptools>
- Bizior-Ociepa R., Jaros V. (2002). Najpopularniejsze imiona żeńskie i męskie nadane w Częstochowie w 2000 roku, „Prace Naukowe. Filologia Polska. Językoznawstwo” nr 4, s. 131–138.
- Bloothoof G., Groot L. (2008). Name Clustering on the Basis of Parental Preferences, „Names” 56, s. 111–163.
- Bloothoof G., Mandemakers K. (2011). Exploring Co-variation in the (Historical) Dutch Civil Registration, [w:] *Els noms en la vida quotidiana. Actes del XXIV Congrés Internacional d'ICOS sobre Ciències Onomàstiques*. Annex. Secció 3, ed. J. Tort i Donada, M. Montagut i Montagut. Barcelona, s. 271–282.
- Bloothoof G., Onland D. (2011). Socioeconomic Determinants of First Names, „Names” 59, s. 25–41.
- Bloothoof G., Onland D. (2016). Multiple First Names in the Netherlands (1760–2014), „Names” 64, s. 3–18.
- Bloothoof G., Schraagen M. P. (2011). Name fashion dynamics and social class, <https://dspace.library.uu.nl/handle/1874/355635> [dostęp: 10 października 2018 r.]

- Bohmbach K. G. (2000). Names and Naming in the Biblical World, [w:] Women in Scripture: A Dictionary of Named and Unnamed Women in the Hebrew Bible, the Apocryphal/Deuterocanonical Books, and the New Testament, red. C. Meyers. Boston, s. 33–99.
- Boksański Z., Piotrowski A., Ziółkowski M. (1977). Socjologia języka. Warszawa
- Bondaletov V. D. (1983). Russkaâ onomastika: Učebnoe posobie. Moskva: Prosvešenie.
- Borek H. (1978). Socjolingwistyczne aspekty imiennictwa, „Onomastica” XXIII, s. 163–175.
- Breza E. (1987). Motywy wyboru imion dla dzieci urodzonych w Bydgoszczy w roku 1966 i 1976.
- Bubak J. (1978). Imiona modne, „Onomastica” XXIII, s. 177–182.
- Bubak J. (1983a). Nadawanie imion w Polsce. Aspekt prawny i językowy, „Socjolingwistyka” 5, s. 145–164.
- Bubak J. (1983b). Wykaz imion używanych w Polskiej Rzeczypospolitej Ludowej, „Język Polski” LXIII, s. 72–84.
- Bubak J. (1988). Nadawanie imion w Polsce w świetle przepisów administracyjnych, [w:] Onomastyka w dydaktyce szkolnej i społecznej. Materiały z VI Konferencji Onomastycznej, pod red. E. Homy. Szczecin, s. 173–182.
- Bubak J. (1993). Księga naszych imion. Wrocław.
- Bubak J. (1998). Zmiany w zasobie imion chrzestnych współczesnych Polaków. Najnowsze przemiany nazewniczy, red. E. Jakus-Borkowa, K. Nowik. Opole: Energeia, s. 47–55.
- Bubak J., Mrózek R. (1983). O współczesnych motywach wyboru i popularności imion, „Rocznik Cieszyński” IV–V, s. 151–164.
- Bystroń J. S. (1938). Księga imion w Polsce używanych. Warszawa.
- Carteron A., Jeanmougin M., Leprieur F., Spatharis S. (2012). Assessing the efficiency of clustering algorithms and goodness-of-fit measures using phytoplankton field data, „Ecological Informatics” 9 (2012), s. 64–68
- CBOS (2014). Wieś polska – stereotypy, „Komunikat z badań CBOS” BS/4/2014. Warszawa.

- Cieślíkowa A. (1988). Społeczny i kulturowy aspekt nadawania imion w Polsce, [w:] Onomastyka w dydaktyce szkolnej i społecznej. Materiały z VI Ogólnopolskiej Konferencji Onomastycznej, red. E. Homa, Szczecin, s.193–198.
- Cieślíkowa A. (2013). Oficjalność i wariantywność imion nadawanych w Polsce, „Onomastica” LVII, s. 129–143.
- Cieślíkowa A., Malec M., Rymut K. (red.) (1995–2002). Słownik etymologiczno-motywacyjny staropolskich nazw osobowych. T. I–VII. Kraków.
- Csárdi G., Nepusz T (2006). The igraph software package for complex network research, „InterJournal”, Complex Systems 1695, <https://cran.r-project.org/web/packages/igraph/>
- Czajkowska A. (2007). Zmiana imion i nazwisk. Warszawa.
- Czaplicka M. (2008). Współczesne tendencje w nadawaniu imion dzieciom w Łobżenicy, „Onomastica Slavogermanica” XXVII, s. 179–187.
- Czaplicka-Jedlikowska M. (2008). Współczesne tendencje kulturowe w nadawaniu imion dzieciom na przykładzie Łobżenicy, „Linguistica Bidgostiana”, t. V, s. 96–110.
- Czopek-Kopciuch B., 2013, Frekwencja i geografia imion rzadkich w Polsce, „Onomastica” LVII, s. 71–93.
- Dobrzyński W. (1974). Słowotwórstwo imion w gwarze Adamowa i okolicy, „Zeszyty Naukowe WSP w Opolu. Językoznawstwo” V, s. 19–32.
- Doroszewicz (2016). Imię jako narzędzie wpływu społecznego, [w:] Wpływ społeczny – konteksty i zastosowania badawcze, red. B. Pająk-Patkowska Beata. Poznań. s. 29–44.
- Doroszewicz K. (2013). Z imieniem przez życie. Psychologiczne aspekty imion ludzkich. Warszawa.
- Doroszewicz K., Krzezińska M. (2007). Forma imienia jako przesłanka wnioskowania o osobowości, „Przegląd Psychologiczny” 50, s. 259–272.
- Doroszewicz K., Stanisławiak E. (1999). Oblicza imion: Wprowadzenie do psychologii imion ludzkich. Warszawa.
- Dubisz S. (1995). Rozwój współczesnej polszczyzny, „Przegląd Humanistyczny” 5, s. 69–85.

- Dunaj B. (1988). Frekwencja w interpretacji zmian językowych, „Biuletyn Polskiego Towarzystwa Językoznawczego” XLI, s. 83–89.
- Dyczewski L. (1995). Kultura polska w okresie przemian. Lublin.
- Dziubok A., Linek B. (2011). W Polsce Ludowej (1945–1989), [w:] Historia Górnego Śląska. Polityka, gospodarka i kultura europejskiego regionu, red. J. Bahlcke, D. Gawrecki, R. Kaczmarek. Gliwice, s. 267–286.
- Eder M. (2010). Does Size Matter? Authorship Attribution, Small Samples, Big Problem, [w:] Digital Humanities 2010, DH 2010, Conference Abstracts, King's College London, London, England, UK, July 7-10, 2010. London, s. 132–5.
- Eder M. (2014). Metody ścisłe w literaturoznawstwie i pułapki pozornego obiektywizmu – przykład stylometrii, „Teksty Drugie” 2, s. 90–105.
- Eder M. (2017). Visualization in stylometry: Cluster analysis using networks, „Digital Scholarship in the Humanities”, Vol. 32, Iss. 1, s. 50–64.
- Eder M., Kestemont M., Rybicki J. (2013). Stylometry with R: a suite of tools, [w:] Digital Humanities 2013: Conference Abstracts. University of Nebraska–Lincoln, s. 487–89.
- Eshel A. (2001). On the Frequency Distribution of First Names, „Names” 49, s. 55–60.
- Evans C. K. (2007). The Tsunami Curve and Popular Culture Influences on Given Names, „Names” 55, s. 335–342.
- Evert S., Baroni M. (2007). zipfR: Word frequency distributions in R, [w:] Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Sessions. Prague, s. 29–32. (R package version 0.6-6 of 2012-04-03).
- Flizak S. (1937). Moda w nadawaniu imion chrzestnych u górali pod Gorcami, „Lud” XXXV, s. 55–59.
- Folfasiński S. (1947–1949). Imiennictwo w Częstochowie [w latach 1946–1948], „Życie Częstochowy” nr 9–10 (1947), nr 24–25 (1948), nr 11–12 (1949).
- Fros H., Sowa F. (1997–2007). Księga imion i świętych. T. I–VI. Kraków.
- Gajda S. (1973). Socjologia imion na przykładzie Opola i okolic, „Studia Śląskie. Seria nowa”, t. XXIII, s. 267–296.

- Galbi, D. A. (2001). A new account of personalization and effective communication. <http://www.galbithink.org/pers.pdf> [dostęp: 28 stycznia 2019 r.].
- Galili T. (2015). dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*. DOI: 10.1093/bioinformatics/btv428
- Gąsiorek K. (1997). *Od słowa do tekstu. O języku i stylach Janusza Korczaka*. Kraków.
- Gerhards J. (2005). *The Name Game. Cultural modernization and first names*. New Brunswick–London.
- Gerritzen D. (1998/99). Changes in the naming patterns for girls and boys in the Netherlands against the cultural background (XXth century), „*Onoma*” 34, s. 181-195
- Gołębiowska T. (1971). *Antroponimia Orawy*. Kraków.
- Gorzelał G. (2007). *Rozwój – region – polityka*, [w:] *Rozwój, region, przestrzeń*, red. G. Gorzelał, A. Tucholska. Warszawa, s. 179–214.
- Gralewski M. (1877). O imionach używanych u ludu polskiego w ziemi łęczyckiej, „*Zbiór Wiadomości do Antropologii Krajowej*” I, s. 97–100.
- Guiraud P. (1954). *Les caractères statistiques du vocabulaire*. Paris.
- Guo J.-Z., Chen Q.-H., Wang Y.-G. (2011). Statistical distribution of Chinese names, „*Chinese Physics B*” 20, 118901, s. 1–7.
- Gureckis T. M., Goldstone R. L. (2009). How You Named Your Child: Understanding the Relationship Between Individual Decision Making and Collective Outcomes, „*Topics in Cognitive Science*” 1, s. 651–674.
- Hahn M. W., Bentley A. (2003). Drift as a mechanism for cultural change: an example from baby names, „*Proceedings of the Royal Society London B. Biology Letters*” 270, s. 120–123.
- Halamska M. (2007). *Przemiany polskiej wielkości*, [w:] *Rozwój, region, przestrzeń*, red. G. Gorzelał, A. Tucholska. Warszawa, s. 235–263.
- Hammerl R., Sambor J. (1990). *Statystyka dla językoznawców*. Warszawa.
- Handke K. (1986/87). O regionalizmie frekwencyjnym, „*Studia Polonistyczne*” XIV–XV, s. 71–76.

- Hoover D. L. (2003). Another Perspective on Vocabulary Richness, „Computers and the Humanities” 37, s. 151–178.
- Hout R., Vermeer A. (2007). Comparing measures of lexical richness, [w:] *Modelling and assessing vocabulary knowledge*, red. H. Daller, J. Milton, J. Treffers-Daller. Cambridge, s. 93–116.
- Hrynicky W. M. (2010). *Nadawanie oraz zmiana imion i nazwisk*. Wrocław.
- Hryniewicz-Adamskich B. (2018). Zastosowanie ilościowo-jakościowej metody statystyki lingwistycznej w badaniach siedemnastowiecznego zasobu chrześcijańskich imion osobowych północnej Rosji, „*Onomastica*” LXII, s. 183–202.
- Huschka D., Gerhards J., Wagner G. G. (2009). Naming Differences in Divided Germany, „*Names*” 57, s. 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.
- Iglewska S. (2015). O imionach dzieci urodzonych w Gnieźnie w latach 2010–2013, [w:] *Funkcje nazw własnych w kulturze i komunikacji*, red. I. Sarnowska-Gieffing, M. Balowski, M. Graf. Poznań.
- Jałowiecki B. (1996). Przestrzeń historyczna, regionalizm, regionalizacja, [w:] *Oblicza polskich regionów*, red. B. Jałowiecki (Seria: *Studia regionalne i lokalne*, nr 50). Warszawa, s. 19–88.
- Jarozuk J. (1983). Imiona chrzestne w parafii Białystok (XVIII w.), [w:] *Warsztat współczesnego onomasty*, [w:] *Materiały z III Ogólnopolskiej Konferencji Onomastycznej*, red. D. Kopertowska. Kielce.
- Jarvis S. (2002). Short texts, best-fitting curves and new measures of lexical diversity, „*Language Testing*” 19, s. 57–84.
- Jóźwiak D. (2017). Imiona chrzestne mieszkańców parafii Gdeszyn w świetle ksiąg metrykalnych z lat 1930–1970, „*Rocznik Lubelskiego Towarzystwa Genealogicznego*” VIII, s. 111–128.

- Kałuża-Kopias D. (2014). Atrakcyjność migracyjna wielkich miast – stan obecny i perspektywy, „Problemy Polityki Społecznej. Studia i Dyskusje” t. 4, nr 27, s. 41–54.
- Kamińska M. (1986). Nazwy osobowe w najstarszych księgach metrykalnych parafii Łask, „Onomastica” XXX, s. 159–179.
- Kamińska-Szmaj I. (1990). Różnice leksykalne między stylami funkcjonalnymi polszczyzny pisanej. Analiza statystyczna na materiale słownika frekwencyjnego. Wrocław.
- Karaś M. (1969). Antroponimia, czyli nazwy osobowe Przędzela i okolicy w pow. nizańskim, „Profile. Kwartalnik Rzeszowski” 1, s. 25–48.
- Klemensiewicz Z. (1981). Historia języka polskiego. Warszawa.
- Klisiewicz E. (1989). Męskie imiona chrzestne w parafii Rudawa koło Krakowa z lat 1570–1897 (na podstawie ksiąg metrykalnych), „Onomastica” XXXIII, s. 167–181.
- Klisiewicz E. (1989). Żeńskie imiona chrzestne w podkrakowskiej parafii Rudawa z lat 1570–1897 (na podstawie ksiąg metrykalnych), „Rocznik Naukowo-Dydaktyczny WSP w Krakowie. Prace Językoznawcze” 8, z. 168, s. 77–82.
- Kosiński L. (1968). Migracje ludności w Polsce w latach 1950–1960. Warszawa.
- Kowalska D. (2013). Sztuka słowa Mikołaja Reja. Studium stylistycznojęzykowe Psalterza Dawidowego. Łódź.
- Kresa M. (2013). Antroponimia historycznego pogranicza mazowiecko-podlaskiego w XVIII wieku na przykładzie parafii Stoczek w ówczesnym dekanacie kamieńczukowskim: Imiona. Warszawa.
- Kuchta A. (2013). Imiona chrzestne w metrykach chrztów parafii rzymsko-katolickiej p.w. św. Jozafata w Korczówce, „Studia Wschodniosłowiańskie” 13, s. 47–59.
- Lewandowska D. (2008). Struktura statystyczna leksyki informacji prasowych (na przykładzie wiadomości z lat 60. XX w. i współczesnych), „Studia medioznawcze” nr 3 (34), s. 72–83.
- Lewaszkiwicz T., Rzepka W. R. (1988). Zmiana tradycji w zakresie nadawania imion jako efekt integracji społecznej i kulturowej (na przykładzie gminy Słońsk w woj. gorzowskim), [w:] Onomastyka w dydaktyce szkolnej i społecznej. Materiały

z VI Ogólnopolskiej Konferencji Onomastycznej, red. E. Homa, Szczecin, s. 215–223.

Li W. (2012). Analyses of Baby Name Popularity Distribution in U.S. for the Last 131 Years, „Complexity” 18, s. 44–50.

Lieberson S. (2000). A Matter of Taste. How Names, Fashions, and Culture Change. New Haven.

Lieberson S., Bell E. O. (1992). Children's first names: An empirical study of social taste, „American Journal of Sociology” 98(3), s. 511–554.

Lieberson S., Lynn F. B. (2003). Popularity as a taste: an application to the naming process, „Onoma” 38, s. 235–276.

López-Escobedo F., Méndez-Cruz C.-F., Sierra G., Solórzano-Soto J. (2013). Analysis of Stylometric Variables in Long and Short Texts, „Procedia – Social and Behavioral Sciences” 95, s. 604–611

Lubaś W. (1978). Typ kontaktu językowego a wariantywność nazw własnych we współczesnej polszczyźnie, „Prace Językoznawcze” 4. Katowice, s. 7–15.

Lubaś W. (2013). Metoda socjolingwistyczna w badaniach onomastycznych, [w:] W. Lubaś, Studia socjolingwistyczne. Opole, s. 67–75.

Luyckx K., Daelemans W. (2011). The effect of author set size and data size in authorship attribution, „Literary and Linguistic Computing” 26(1), s. 35–55.

Machowska-Szewczyk M., Sompolska-Rzechuła A. (2012). Ocena stabilności wyników klasyfikacji województw polski pod względem poziomu życia ludności, „Metody ilościowe w badaniach ekonomicznych” XIII/3, s. 159–168.

Magda-Czekaj M. (2000). Imiona mieszkańców Olesna na Śląsku Opolskim i ich socjologiczne uwarunkowanie od XVI do 1. połowy XX wieku, „Onomastica” XLV, s. 191–210.

Malec M. (1998a). Imiona, [w:] Polskie nazwy własne. Encyklopedia. Red. E. Rzetelska-Feleszko. Warszawa-Kraków, s. 97–118.

Malec M. (1998b). Zjawisko geminacji w antroponimii polskiej, [w:] Najnowsze przemiany nazewnicze, red. E. Jakus-Borkowa, K. Nowik. Opole: Energeia, s. 299–305.

- Malec M. (2007). Stopnie otwartości systemu imion w Polsce, [w:] Nowe nazwy własne – nowe tendencje badawcze. Red. A. Cieślikowa, B. Czopek-Kopciuch, K. Skowronek. Kraków, s. 125–137.
- Malvern D., Richards B., Chipere N., Durán P. (2004). *Lexical Diversity and Language Development: Quantification and Assessment*. New York.
- Maślach A. (2007). Imiennictwo wsi galicyjskiej w XIX w. (na podstawie metryk z Zarzecza, pow. nizański z l. 1828–1864), „Rocznik Przemyski” XLIII (z. 3, Literatura i język), s. 125–147.
- Mateos P., Tucker K. (2008). Forenames and Surnames in Spain, „Names” 56, s. 165–184.
- McGarigal K., Cushman S. A., Stafford S. (2000). *Multivariate Statistics for Wildlife and Ecology Research*. New York.
- Mehri A., Jamaati M. (2017). Variation of Zipf's exponent in one hundred live languages: A study of the Holy Bible translations, „Physics Letters A”, vol. 381, s. 2470–2477
- Meyers C. (red.) (2000). *Women in Scripture. A Dictionary of Named and Unnamed Women in the Hebrew Bible, the Apocryphal/Deuterocanonical Books, and the New Testament*. New York.
- Milewski T. (1963). Imiona osobowe jako zwierciadło kultury, [w:] *Studia linguistica in honorem Thaddaei Lehr-Spławińskiego*, red. T. Milewski, J. Safarewicz, F. Sławski. Warszawa, s. 39–46.
- Miranda-García A., Calle-Martín J. (2005). Yule's Characteristic K Revisited, „Language Resources and Evaluation” 39(4), s. 287–294
- Mistrík J. (1967). Matematiko-statističeskie metody v stilistike, „Voprosy Ťazykoznaniâ” 16(3), s. 42–52.
- Mordań M. (2017). Imiona chrzestne bielskich prawosławnych i katolików na przełomie XIX i XX wieku, „Studia Wschodniosłowiańskie” 17, s. 247–267.
- Moskovich I. (2016). Lexical richness in modern women writers: Evidence from the Corpus of History English Text, „Revista Canaria de Estudios Ingleses” nr 72, s. 111–128.
- Niewyk D. L. (2000). *The Columbia Guide to the Holocaust*. New York.

- Nowik, K. (1998). Zmiany frekwencyjne w zasobie imion w Polsce powojennej, [w:] Najnowsze przemiany nazewnictwa, red. E. Jakus-Borkowa, K. Nowik. Opole, s. 57–71.
- Ostroróg-Sadowski J. (1902). O imieniu i nazwisku. Studium prawne. Warszawa.
- Parzniewska J. (1996). Imiona nadawane w Krakowie w 1992 roku, „Onomastica” XLI, s. 139–151.
- Pawłowski A. (2003). O problemie atrybucji tekstu w lingwistyce kwantytatywnej (na przykładzie tekstów Romaina Gary), [w:] Prace językoznawcze dedykowane Profesor Jadwidze Sambor, red. J. Linde-Usiekniewicz, R. Huszcza. Warszawa, s. 169–190.
- Piotrowski A., Ziółkowski M. (1976). Zróznicowanie językowe a struktura społeczna. Warszawa.
- Plichta R. (1998). Imiona chrzestne w księgach metrykalnych katedry łowickiej w latach 1951–1955 i 1991–1995, „Onomastica” XLIII, s. 117–139.
- Plit J. (2015). Regionalizacja współczesnych krajobrazów historyczno-kulturowych Polski, „Prace Komisji Krajobrazu Kulturowego
- Pomorski M., Krawczyk M. J., Kułakowski K., Kwapien J., Ausloos M. (2016). Inferring cultural regions from correlation networks of given baby names, „Physica A” 445, s. 169–175.
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna. [<http://www.R-project.org/>]
- Rejman S. (2017). Imiona nadawane dzieciom w rzeszowskiej parafii farnej na początku XX wieku, „Przeszłość Demograficzna Polski” 39, s. 253–281.
- Romanik P., Kiper D. (2016). Imiona chrzestne nadawane w wybranych miejscowościach rzymskokatolickiej parafii Pruchnik w okresie przedautonomicznym, „Archiwa, Biblioteki i Muzea Kościelne” 106, s. 235–274.
- Ruczkowski P. (2010). Imię i nazwisko w prawie polskim. Kielce.
- Rudnicka-Fira E. (1996). Nazewnictwo, [w:] Język polski czasu drugiej wojny światowej, red. I. Bajerowa. Warszawa, s. 260–270.

- Rudnicka-Fira E. (2011). Imiona mieszkańców miast Górnego Śląska i Zagłębia Dąbrowskiego w perspektywie socjolingwistycznej, „Onomastica Slavogermanica” XXX, s. 197–207.
- Rudnicka-Fira E. (2013). Imiennictwo krakowian od XVI do XVIII wieku na tle historii i kultury. Kraków.
- Ruszkowski M. (2003). Liczebność próby w statystyczno-składniowych badaniach stylu, „Język Polski” LXXXIII/3, s. 174–180.
- Ruszkowski M. (2005). Polskie badania z zakresu statystyki językoznawczej. Próba syntezy, „Respectus Philologicus” nr 7 (12), s. 49–56.
- Rybicki J. (2008). Does Size Matter? A Re-examination of a Time-proven Method, [w:] Digital Humanities 2008: Book of Abstracts. University of Oulu, Finland, s. 184.
- Rybicki J. (2011). Alma Cardell Curtin and Jeremiah Curtin: the translator’s wife’s stylistic fingerprint, [w:] Digital Humanities 2011: Conference Abstracts. Stanford, s. 308–11.
- Rybicki J., Heydel M. (2013). The stylistics and stylometry of collaborative translation: Woolf’s Night and Day in Polish, „Literary and Linguistic Computing”, 28(4), s. 708–717.
- Ryguła E. M. (2011). Imiennictwo w Kobiórze w XVIII wieku, „Wieki Stare i Nowe” 3(8), s. 104–127.
- Rymut K. (1995). Słownik imion współcześnie w Polsce używanych. Kraków.
- Sambor J. (1969). Badania statystyczne nad słownictwem (na materiale „Pana Tadeusza”). Wrocław.
- Sambor J. (1972). Słowa i liczby. Zagadnienia językoznawstwa statystycznego. Wrocław.
- Saraçlı S., Doğan N., Doğan İ. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation, „Journal of Inequalities and Applications” 2013: 203.
- Skowronek K. (2001). Współczesne nazwisko polskie. Studium statystyczno-kognitywne. Kraków.
- Skowronek K. (2013). Imiona „wielkomiejskie” w latach 1995–2010 w perspektywie statystyczno-onomastycznej i społeczno-kulturowej, „Onomastica” LVII, s. 95–127.

- Skulina T. (1988). Imiona kobiet w Wielkopolsce XVII i XVIII wieku, „Slavia Occidentalis” XLV, s. 47–60.
- Sobczyk M. (1998). Statystyka. Podstawy teoretyczne. Przykłady – zadania. Lublin.
- Spyt A. (1973). O imionach podwójnych typu *Stanisław Marian* we wsiach zachodniej części powiatu chrzanowskiego, „Onomastica” XVIII, s. 245–255.
- Surma G. (1986). Imiona mieszkańców w gminie Debrzno w woj. słupskim, „Onomastica” XXX, s. 181–211.
- Swoboda P. (2013). Imiona częste w Polsce w latach 1995–2010 oraz ich zróżnicowanie w czasie i przestrzeni, „Onomastica” LVII, s. 19–69.
- Swoboda P. (2015). Regionalne zróżnicowanie ilościowe imiennictwa w Polsce, „Socjolingwistyka” 29, s. 21–37.
- Szulowska W. (2004). Dawna antroponimia Mazowsza (XV–XVII w.). Olsztyn.
- Szymańska D. (2002). Niektóre zagadnienia urbanizacji w Polsce w drugiej połowie XX wieku, [w:] Przemiany bazy ekonomicznej i struktury przestrzennej miast, red. J. Słodczyk. Opole, s. 53–69.
- Tan P.-N., Steinbach M., Kumar V. (2005). Introduction to Data Mining. Boston.
- Taszycki W. (1925). Najdawniejsze polskie imiona osobowe. Kraków.
- Taszycki W. (red.) (1965–1987). Słownik staropolskich nazw osobowych. T. I–VII. Wrocław.
- Tomecka A. (1994). Imiona chrzestne w księgach metrykalnych Nowej Brzeźnicy k. Częstochowy w XVII i XVIII wieku, „Onomastica” XXXIX, s. 157–178.
- Tucker D. K. (2001). Distribution of Forenames, Surnames, and Forename–Surname Pairs in the United States, „Names” 49/2, s. 69–96.
- Tucker D. K. (2002). Distribution of Forenames, Surnames, and Forename–Surname Pairs in Canada, „Names” 50/2, s. 105–132.
- Tucker K. (2004). The forenames and surnames from the GB 1998 Electoral Roll compared with those from the UK 1881 census, „Nomina” 27, s. 5–40.
- Tucker K. (2009). Increased Competition and Reduced Popularity: US Given Name Trends of the Twentieth and Early Twenty-First Centuries, „Names” 57, s. 52–62

- Tuldava J. (2005). Stylistics, author identification, [w:] *Quantitative Linguistik: Ein internationales Handbuch / Quantitative Linguistics: An International Handbook*, red. R. Köhler, G. Altmann, R.G. Piotrowski. Berlin, s. 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, s. 323–352.
- Twenge J. M., Abebe E. M., Campbell W. K. (2010). Fitting In or Standing Out: Trends in American Parents’ Choices for Children’s Names, 1880–2007, „*Social Psychological and Personality Science*” I(I), s. 19–25.
- Twenge J. M., Dawson L., Campbell W. K. (2016). Still standing out: children’s names in the United States during the Great Recession and correlations with economic indicators, „*Journal of Applied Psychology*” 46, s. 663–670.
- Tylec G. (2013). *Imię i nazwisko osoby fizycznej oraz ich ochrona w polskim prawie cywilnym*. Lublin.
- Umińska E. (1983). Imiona metrykalne mieszkańców okolic Działoszyna, „*Onomastica*” XXVIII, s. 133–163.
- Umińska-Tytoń E. (1987). Popularne imiona w Łodzi, „*Acta Universitatis Lodzianensis. Folia Linguistica*” 15, s. 61–91.
- Umińska-Tytoń E. (1988). Imiona Łodzian urodzonych podczas okupacji hitlerowskiej, „*Onomastica*” XXXII, s. 121–140.
- Walczak M. (1980). Imiona chrzestne w parafii Włostowice w Puławach w wiekach XVIII–XX, „*Język Polski*” LX, s. 163–169.
- Warchala J. (2003): *Kategoria potoczności w języku*. Katowice.
- Wilson S. (1998). *The Means of Naming: A Social and Cultural History of Personal Naming in Western Europe*. London.
- Włodarski Ł. (2015). Imiona chrzestne w parafii rzymskokatolickiej pw. św. Stanisława B. M. w Drobinie w latach 1826–1913, „*Archiwa, Biblioteki i Muzea Kościelne*” 104, s. 349–362.
- Wolnicz-Pawłowska E., Szulowska W. (1998). *Antroponimia polska na kresach południowo-wschodnich. XV–XIX w.* Warszawa.

Zaręba A. (1957), (1959). Polskie imiona ludowe, „Onomastica” III, s. 129–178, V, s. 419–446.

Zielińska A. (2012). Imiona chrzestne wśród katolików i ewangelików w Toruniu w XIX i na początku XX wieku, „Rocznik Toruński” 39, s. 161–179.