

# **Die Verlustlisten des Ersten Weltkriegs als historisches namengeographisches Korpus**

## **1. Einleitung**

Für die großflächige, deutschlandweite Kartierung und Analyse der Familiennamen (FamilienN, im Folgenden steht „N“ für Name(n)) stellt bis heute die nach Postleitzahlen gegliederte Datenbank der Telefonanschlüsse, wie sie im Telefonbuch der Deutschen Telekom von 2005 verzeichnet sind die Datenbasis des Projekts 'Deutscher Familiennamenatlas' dar. Im Folgenden werden wir eine neue umfangreiche Datenbasis der deutschen FamilienN vorstellen, die die bisherige um eine historische Dimension erweitern wird. Es handelt sich um die Datenbank der Verlustlisten der Kriegsteilnehmer des Ersten Weltkriegs, die mit ihren rund 7,7 Millionen Namenträgern aus allen Gebieten des damaligen Deutschen Reichs die gesamte FamilienN-Landschaft des ausgehenden 19. Jh. umfassen. Mit dieser neuen Datenbasis wird es möglich sein, die heutige FamilienN-Landschaft mit derjenigen vor 130 Jahren zu vergleichen und damit zum ersten Mal die regionale Festigkeit bzw. Veränderung der FamilienN zu erforschen. Unberührt von den großen europäischen Migrationswellen des 20. Jh. und erst am Beginn der durch die Industrialisierung beschleunigten Urbanisierung, repräsentieren die Verlustlisten die Namenverteilung einer im Vergleich zu heute deutlich weniger mobilen, überwiegend ländlichen Gesellschaft. Über diese neue Kartierungsgrundlage hinaus leisten die Verlustlisten wertvolle Dienste bei der Etymologisierung, bei sprachgeschichtlichen, dialektologischen oder graphematischen Fragestellungen sowie für die Migrationsforschung.

Der Beitrag ist wie folgt gegliedert: Abschnitt 2 erläutert den Aufbau der Verlustlisten, die Aufbereitung in einer Datenbank und gibt globale Einblicke in das onymische Material. Abschnitt 3 präsentiert Kartierungen dieses historischen Materials für exemplarische FamilienN. In Abschnitt 4 schließlich wird eine Isonymie-Analyse vorgestellt, mit deren Hilfe auf statistischem Wege die historischen FamilienN -Areale ermittelt und mit den heutigen sowie den Grenzen der traditionellen Dialekte kontrastiert wird.

## **2. Verlustlisten als Datenquelle, Datenaufbereitung**

In den Verlustlisten wurden von 1914 bis 1919 kontinuierlich, i.d.R. täglich, die Verluste der Streitkräfte des Deutschen Reichs veröffentlicht. Sie umfassen 2535 Ausgaben mit 31202 dreispaltigen Seiten, auf denen die Verluste an Soldatenleben, sei es durch Tod, Krankheit, Verwundung, Gefangenschaft o.Ä. verzeichnet sind.<sup>1</sup> Die Personendaten enthalten u.a. folgende Angaben: FamilienN, Rufname (!), Geburtsort, Art des 'Verlusts', teilweise auch Geburtsjahr, Sterbedatum, Dienstgrad oder Truppenteil. Die Zahl der in den Verlustlisten publizierten Personen wird auf über neun Millionen geschätzt, deren

---

<sup>1</sup> Vgl. 'Verlustlisten Erster Weltkrieg', < [http://wiki-de.genealogy.net/Verlustlisten\\_Erster\\_Weltkrieg](http://wiki-de.genealogy.net/Verlustlisten_Erster_Weltkrieg)>, letzter Aufruf 23.5.2019



wurden und zum Kriegsdienst einberufen wurden.<sup>6</sup> In den Verlustlisten kann es vorkommen, dass Personen mehrfach aufgeführt werden, z.B. wenn ein Soldat zunächst als vermisst oder verwundet und in einer späteren Ausgabe der Verlustlisten als gefallen gemeldet wird oder wenn die Schreibweisen von Namen oder Ortschaften in einem späteren Neueintrag korrigiert werden (z.B. *Aakmann (nicht Ackermann)*, *Abenard Gerhard [nicht Johann]*, *Schorstedt (nicht Schorrstedt)*). Es ist im Moment nicht absehbar, wie stark dies die Güte der Daten verfälscht; allerdings dürften diese Doppelungen angesichts der Menge der Daten wohl nicht ins Gewicht fallen.

Um die Aussagekräftigkeit der Verlustlisten zu bestimmen, werden in Tab. 1 die Kennzahlen verschiedener Datenquellen der FamilienN miteinander kontrastiert. In Relation zu einer damaligen Gesamtbevölkerung von 49,2 Millionen (1890) repräsentieren die Verlustlisten ungefähr 15 % dieser Gesamtheit. Zum Vergleich: Die 26 Millionen Einträge im Telefonbuch von 1996, der bis heute umfassendsten Quelle der FamilienN, entsprechen 31,7 % der Gesamtbevölkerung, also ca. doppelt so viel als die Verlustlisten. Demgegenüber repräsentieren die Daten des Reichstelefonbuchs von 1942, der einzigen verfügbaren Datenquelle für bisherige historische Kartierungen (vgl. z.B. Udolph 2011), lediglich 3,3 % der Bevölkerung. In Ermangelung einer alternativen historischen und digital verfügbaren Datenquelle können die Verlustlisten damit beanspruchen, die umfangreichste und flächendeckendste namenkundlichen Datenbasis zu sein, die es zum ersten Mal erlaubt, die Stabilität und Veränderung der deutschen FamilienN -Landschaft detailliert zu untersuchen. Der Umstand, dass in den Verlustlisten nur Männer erfasst sind, dürfte für die valide Analyse der FamilienN nicht relevant sein.

Tab. 1: Häufigkeitsangaben verschiedener Datensätze der FamilienN

	<b>Namenträger</b>	<b>Gesamtbevölkerung</b>	<b>Abdeckung</b>
Verlustlisten 1. WK	7,7 Mio.	49,2 Mio. (1890)	15,6%
Reichstelefonbuch 1942	2,6 Mio.	79,4 Mio. (1939)	3,3%
Telefonbuch 1996	26 Mio.	82,0 Mio. (1996)	31,7%

Die 7,7 Millionen Tokens verteilen sich auf 460.000 Types. Im Vergleich zum heutigen FamilienN - Gut mit seinen 860.000 Types (nach Nübling et al. 2012:147), enthalten die Verlustlisten also nur ca. die Hälfte, was sicherlich größtenteils darauf zurückzuführen ist, dass zum einen die damalige Gesellschaft weniger durch Migration geprägt war und zum anderen angenommen werden kann, dass (sehr) seltene Namen und Hapaxlegomena in den Verlustlisten in einem geringeren Ausmaß vertreten sind. Dazu passt auch, dass in den Verlustlisten die häufigsten FamilienN mehr Namenträger auf sich vereinigen, als dies heute der Fall ist: Während in den Verlustlisten nämlich die häufigsten 5000

---

<sup>6</sup> Die Zahl der außerhalb des Deutschen Reichs geborenen Soldaten sind natürlich verhältnismäßig gering, so dass in diesen Gebieten bereits verhältnismäßig wenige registrierte Soldaten mit einem einzelnen Namen mitunter enorm hohe relative Häufigkeiten (gemessen an allen dort registrierten Soldaten) ergeben können, vgl. etwa Abb. 8. Die Datenpunkte außerhalb der Grenzen des Deutschen Reichs sind also mit einer gewissen Vorsicht zu behandeln, ihre Größe kann keineswegs in allen Fällen direkt mit den restlichen Punkten verglichen werden. Gleichwohl weisen sie auf das Vorhandensein der jeweiligen Namen am Ort hin, weswegen sie nicht aus den Daten entfernt wurden.

FamilienN 54,3 % der Namenträger repräsentieren, sind es im den Telefonbuchdaten von 1996 lediglich 47,4 %. Dieser Unterschied von 7 % deutet also darauf hin, dass die häufigsten FN früher noch häufiger vorkamen bzw. umgekehrt, dass sich die heutige Namenlandschaft demnach diversifizierter darstellt.

Die folgende Tab. 2 gibt einen ersten Überblick über die Rangfolge der 40 häufigsten Namen kontrastiert mit dem heutigen Zustand (Telefonbuchdaten von 1996). In der Spalte der Rangänderungen deutet sich bereits die relative historische Stabilität der FamilienN an. Für die meisten FamilienN lassen sich nur geringfügige bis gar keine Veränderungen feststellen (positive Zahlen geben an, dass der Name heute höhere Ränge einnimmt, negative Zahlen geben an, dass der Name heute auf niedrigeren Rängen zu finden ist). Darüber hinaus können aber hier bereits Verschiebungen für einzelne FamilienN beobachtet werden, die jeweils gesonderte Analysen verdient: Während für den Rückgang für Lehman (-13 Ränge) oder *Möller* (-11) nicht ohne weiteres Gründe angegeben werden können, so kann beispielsweise die Abnahme von *Weiß* (-14) durch die neu aufgekommene Konkurrenz durch die Schreibvariante *Weiss* erklärt werden, während der Rückgang von *Schulze* (-17) in Zusammenhang mit der Zunahme von apokopiertem *Schulz* zu sehen ist. Generell lässt sich jedoch sagen, dass es während des letzten Jahrhunderts zumindest bei den häufigsten FamilienN zu keinen gravierenden Häufigkeitsverschiebungen gekommen ist. Allerdings kann diese pauschale Beobachtung erst durch (noch zu unternehmende) Detailuntersuchungen bestätigt werden. Es ist durchaus nicht unplausibel, dass sich für einzelne Namen, Namenkomplexe oder phonologisch-graphematische Strukturen historische Verschiebungen ereignet haben.

Tab. 2: Häufigkeitsangaben verschiedener Datensätze der FamilienN

Name	Anzahl 1890	Anzahl 1996	Rang 1890	Rang 1996	Rang-änderung	relative Häufigkeit 1890 [%]	relative Häufigkeit 1996 [%]
<i>Müller</i>	80845	268675	1	1	0	1.0826 %	0.9099 %
<i>Schmidt</i>	60623	195890	2	2	0	0.8118 %	0.6634 %
<i>Meyer</i>	31510	89008	3	5	-2	0.4220 %	0.3014 %
<i>Schneider</i>	30341	116550	4	3	1	0.4063 %	0.3947 %
<i>Schulz</i>	27707	77825	5	8	-3	0.3710 %	0.2636 %
<i>Fischer</i>	27031	100687	6	4	2	0.3620 %	0.3410 %
<i>Weber</i>	22071	86987	7	6	1	0.2956 %	0.2946 %
<i>Hoffmann</i>	21134	73599	8	10	-2	0.2830 %	0.2492 %
<i>Becker</i>	20901	76477	9	9	0	0.2799 %	0.2590 %
<i>Wagner</i>	19996	79215	10	7	3	0.2678 %	0.2683 %
<i>Richter</i>	18606	59288	11	14	-3	0.2492 %	0.2008 %
<i>Koch</i>	17751	61532	12	12	0	0.2377 %	0.2084 %
<i>Schröder</i>	16711	52939	13	16	-3	0.2238 %	0.1793 %
<i>Schäfer</i>	16656	62044	14	11	3	0.2230 %	0.2101 %
<i>Bauer</i>	15674	60190	15	13	2	0.2099 %	0.2038 %
<i>Krüger</i>	15552	44218	16	20	-4	0.2083 %	0.1497 %
<i>Schulze</i>	15243	36989	17	34	-17	0.2041 %	0.1252 %
<i>Neumann</i>	14474	49159	18	18	0	0.1938 %	0.1665 %
<i>Lehmann</i>	14394	37404	19	32	-13	0.1928 %	0.1267 %
<i>Lange</i>	14101	41703	20	24	-4	0.1888 %	0.1412 %
<i>Klein</i>	13883	56022	21	15	6	0.1859 %	0.1897 %
<i>Wolf</i>	13404	51739	22	17	5	0.1794 %	0.1752 %
<i>Krause</i>	13191	40555	23	28	-5	0.1766 %	0.1373 %
<i>Zimmermann</i>	12193	43900	24	21	3	0.1632 %	0.1486 %
<i>Schwarz</i>	11532	45466	25	19	6	0.1544 %	0.1539 %
<i>Werner</i>	11336	39926	26	30	-4	0.1518 %	0.1352 %

<i>Hartmann</i>	11219	41164	27	26	1	0.1502 %	0.1394 %
<i>Schmitz</i>	10887	42075	28	23	5	0.1457 %	0.1424 %
<i>Hofmann</i>	10730	41569	29	25	4	0.1436 %	0.1407 %
<i>Meier</i>	10693	38779	30	31	-1	0.1431 %	0.1313 %
<i>Braun</i>	10552	43168	31	22	9	0.1413 %	0.1461 %
<i>Köhler</i>	10396	34810	32	35	-3	0.1392 %	0.1178 %
<i>Weiß</i>	10196	30302	33	47	-14	0.1365 %	0.1026 %
<i>Möller</i>	9794	31118	34	45	-11	0.1311 %	0.1053 %
<i>Walter</i>	9572	33976	35	38	-3	0.1281 %	0.1150 %
<i>Schmitt</i>	9063	41002	36	27	9	0.1213 %	0.1388 %
<i>König</i>	8910	33695	37	39	-2	0.1193 %	0.1141 %
<i>Scholz</i>	8800	30826	38	46	-8	0.1178 %	0.1043 %
<i>Kaiser</i>	8591	31908	39	42	-3	0.1150 %	0.1080 %
<i>Herrmann</i>	8534	34420	40	37	3	0.1142 %	0.1165 %

### 3. Erste Kartierungen: Stabilität und Variabilität der Namenlandschaft

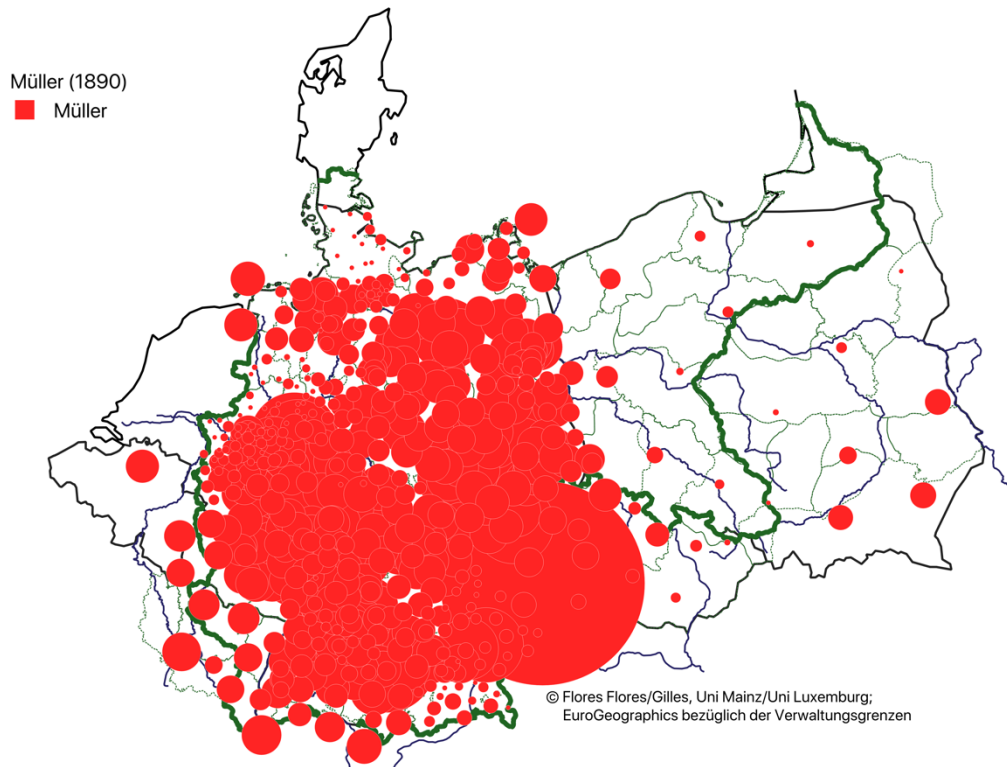
Die Anwendungsmöglichkeiten der aus den Verlustlisten gewonnenen Namendaten sind vielfältig: Am sinnvollsten erscheint uns dabei, die historischen Daten mit rezenten (Telefonbuch-)Daten zu kombinieren, um die FamilienN-Geographie um eine statistisch valide und visualisierbare diachrone Ebene zu ergänzen, die helfen kann unbemerkte Einflüsse jüngerer Entwicklungen in der heutigen Namenlandschaft auszuschließen. Dabei ergänzen sich die jeweiligen Vorteile der beiden Datensätze (größere Abdeckung/Stichprobe der rezenten Daten vs. größere historische Tiefe der Verlustlisten) optimal, so dass sie gerade in der gemeinsamen Betrachtung die besten Ergebnisse liefern können. Für die Kartierung wird hier das GIS-Programm QGIS in Verbindung mit einer PostGIS-Datenbank verwendet, wodurch eine große Auswertungs- und Visualisierungsflexibilität erreicht werden konnte. Um die regionale Vergleichbarkeit zu gewährleisten, werden sowohl die Verlustlisten als auch die heutigen Vergleichsdaten (i.e. Telefonbuch von 1996) auf dieselben dreistelligen Postleitzahlengebiete bezogen. Durch die Kartierung der relativen Häufigkeiten (Namenhäufigkeit in Relation zur Gesamtzahl der Personen pro dreistelligem PLZ-Gebiet) können dann die Namenlandschaften der beiden Zeitschnitte miteinander kontrastiert werden.

Im Folgenden sollen die Möglichkeiten, die sich durch eine solche Verknüpfung rezenter Namendaten mit den Daten der Verlustlisten ergeben, veranschaulicht werden. Im Fokus steht dabei einerseits die grundlegende Stabilität der FamilienN-Landschaft in den letzten rund 150 Jahren – nach Nübling *et al.* (2012:147) sind ca. 85 % der Namen seit Jahrhunderten ortsfest – und andererseits der zugleich bestehende Einfluss der (Nachkriegs-)Migration auf die Verbreitung etlicher FamilienN. Daneben sollen die Beispielkarten auch einen Einblick in die vielfältigen verschiedenen Kartierungs- und Darstellungsmöglichkeiten geben, die moderne GIS wie QGIS bieten. Selbstverständlich können in diesem Beitrag die wissenschaftlichen wie technischen Möglichkeiten allenfalls angerissen werden, in diesem Sinne ist er eher als Anregung zu weiteren Arbeiten zu verstehen: Die Beschreibung der Daten und Kartenbilder steht hier stets im Vordergrund, erste Interpretationen können allenfalls angedeutet werden; alles hier dargestellte kann und sollte weiter verfolgt und tiefer analysiert werden.

### **3.1. Diachrone Stabilität der Namenlandschaft**

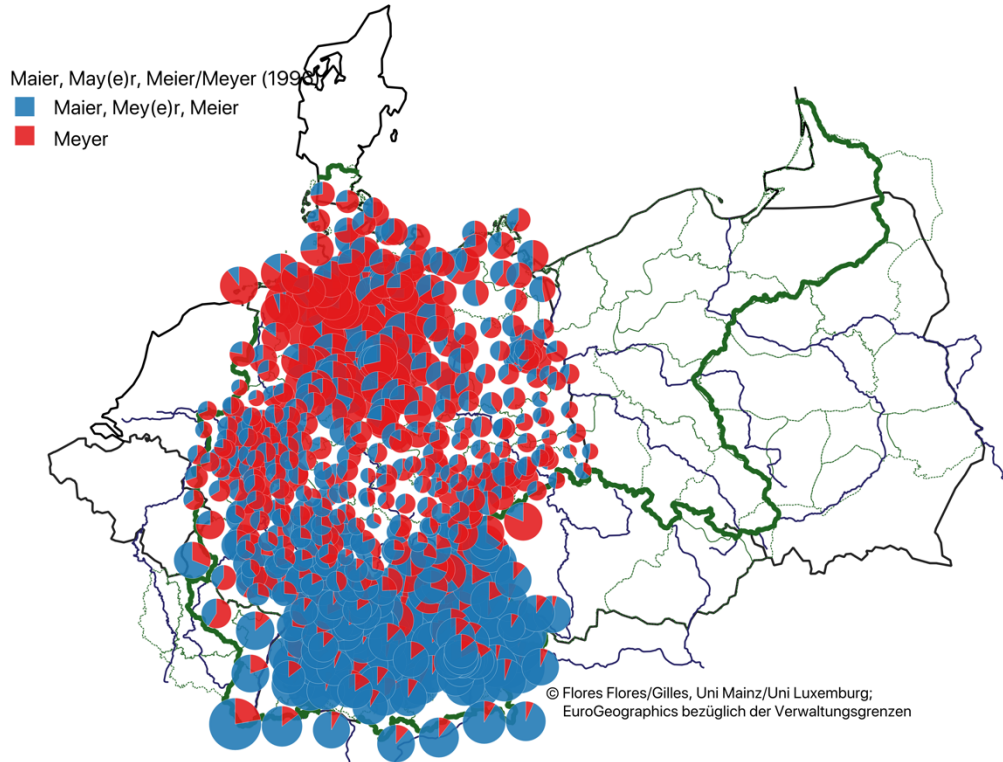
Vergleicht man die Namenlandschaft von 1890 und 1996, so dominiert die vielfach beschriebene diachrone Stabilität die Verbreitung des größten Teils der FamilienN, und zwar sowohl im

großräumigen Rahmen (sehr) häufiger FamilienN wie z.B. Müller (1890 und 1996 Rang 1, vgl. Abb. 2



und

Abb. 3) oder Meyer (1890 Rang 3, 1996 Rang 5, vgl.



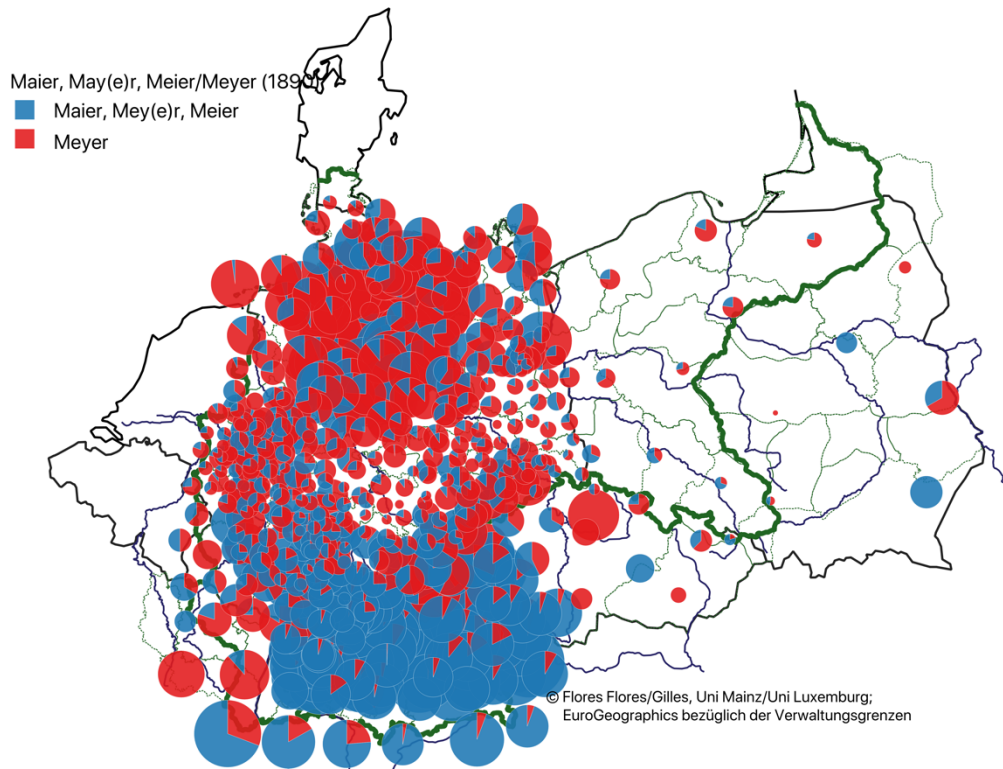
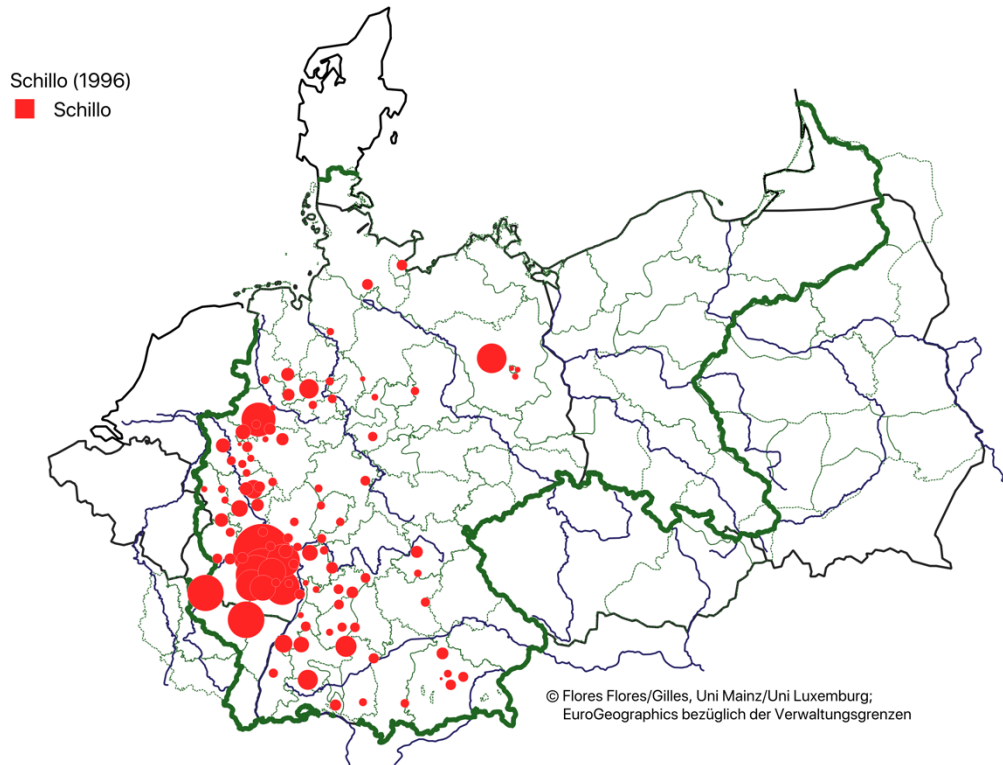


Abb. 5) inkl. seiner Varianten *Maier*, *Mayr*, *Meier*, *Mayer*, als auch im Zusammenhang sehr seltener und nur kleinräumig verbreiteter FamilienN wie z.B. *Schillo* (vgl.





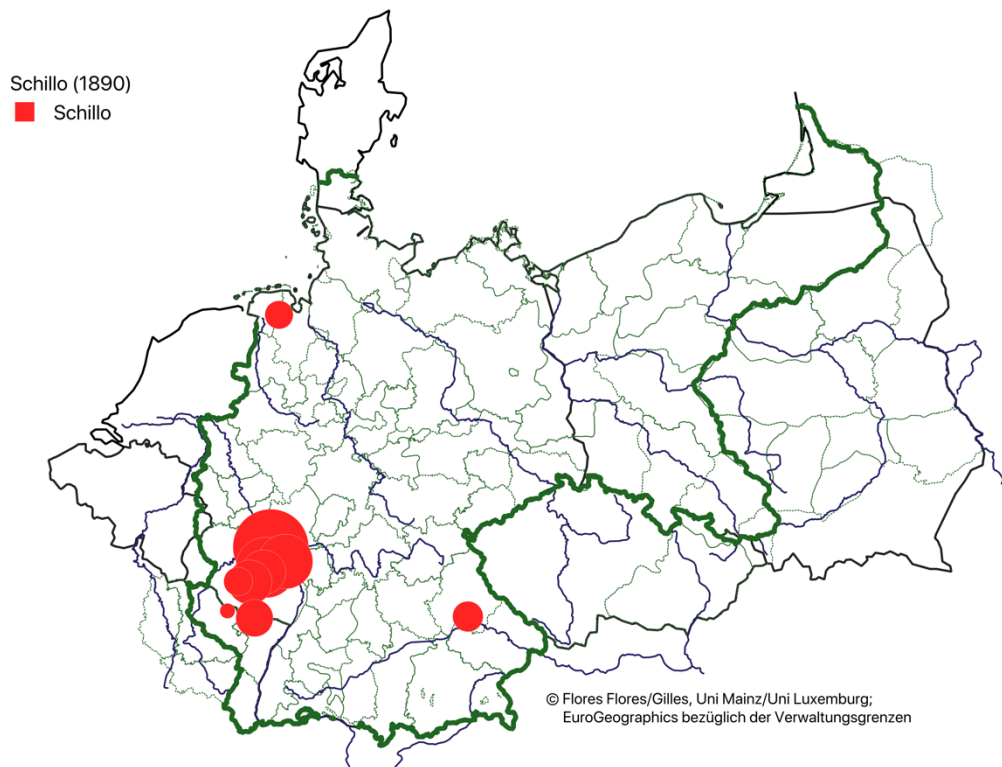


Abb. 7; zur Etymologie dieses saarländischen FamilienN vgl. Heuser/Dräger 2016): Die Verbreitungsmuster sind für beide Zeitschnitte weitgehend identisch. Im Fall von *Müller* lässt sich für 1890 auch seine Verbreitung in Elsass-Lothringen feststellen.

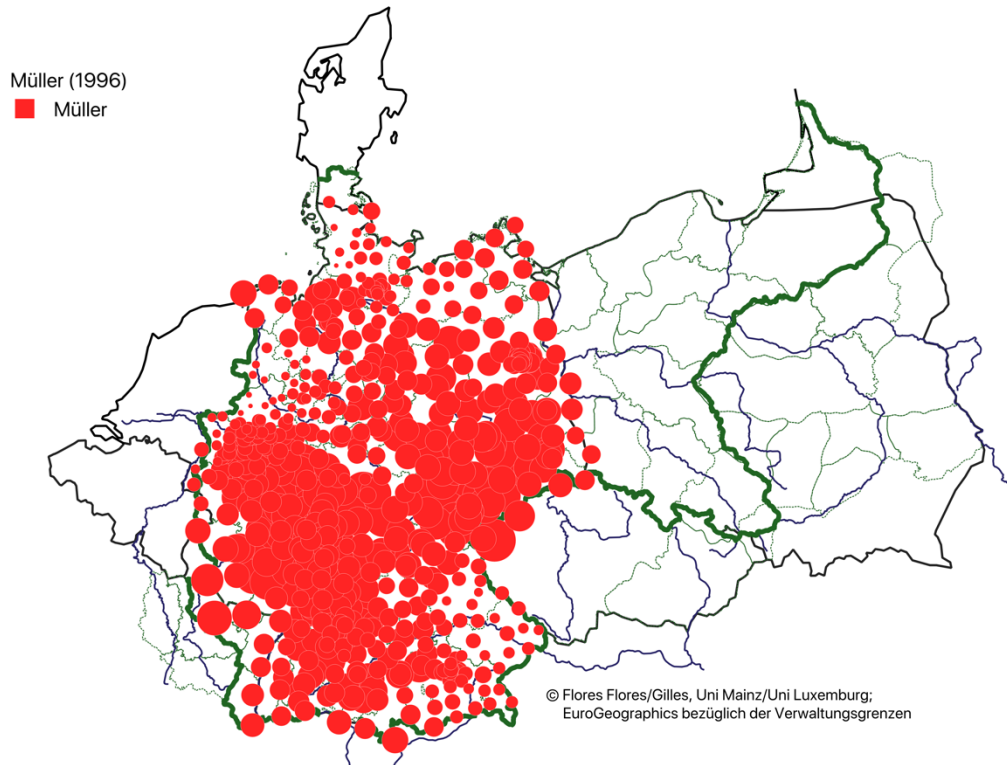
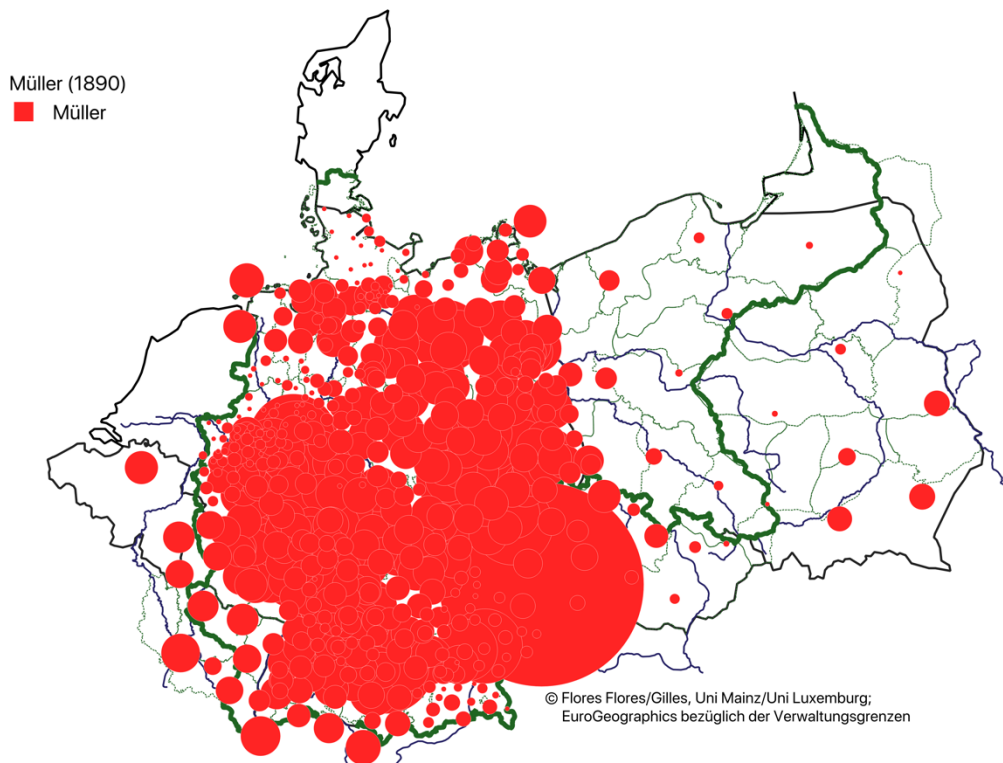


Abb. 2: relative Verbreitung des FamilienN Müller (Max. Symbolgröße: 12, entspricht 25,78%) (1996)<sup>7</sup>

<sup>7</sup> Die Maßstäbe der Kreissymbole sind in Karten zu gleichen Namen hier und stets gleich, um eine optimale Vergleichbarkeit zu gewährleisten; d.h. bei gleicher relativer Häufigkeit erscheinen die Symbole etwa in Abb. 2



und

Abb. 3 gleich. Die Angabe der Symbolgröße bezieht sich (außer bei Abb. 8 und Abb. 9) immer auf die Fläche der Kreisdiagramme.

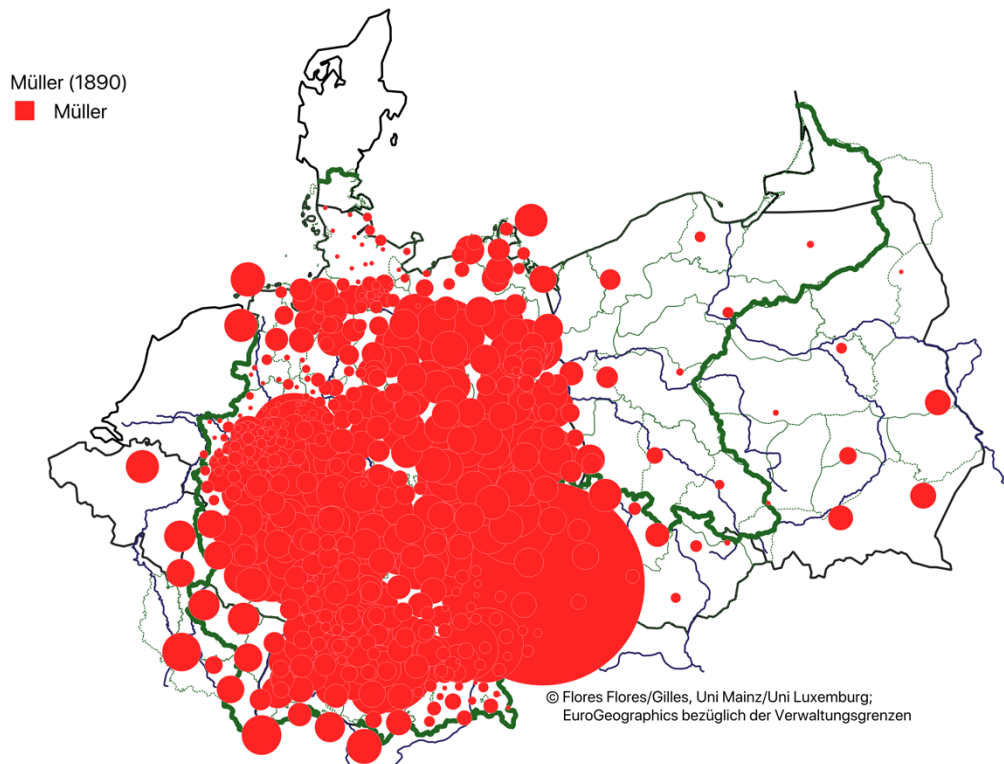


Abb. 3: relative Verbreitung des FamilienN Müller (Max. Symbolgröße: 42, entspricht 90,91%) (1890)

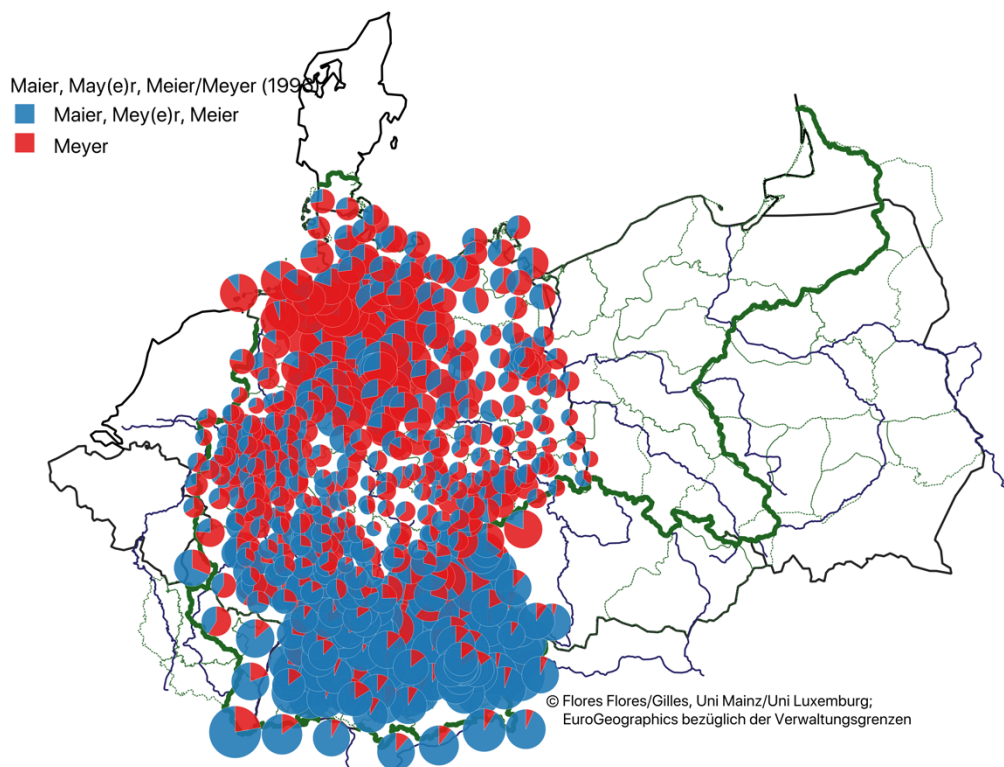


Abb. 4: relative Verbreitung der FamilienN Maier, Mayer, Meier, Mayr vs. Meyer (Max. Symbolgröße: 12, entspricht 24,69%) (1996)

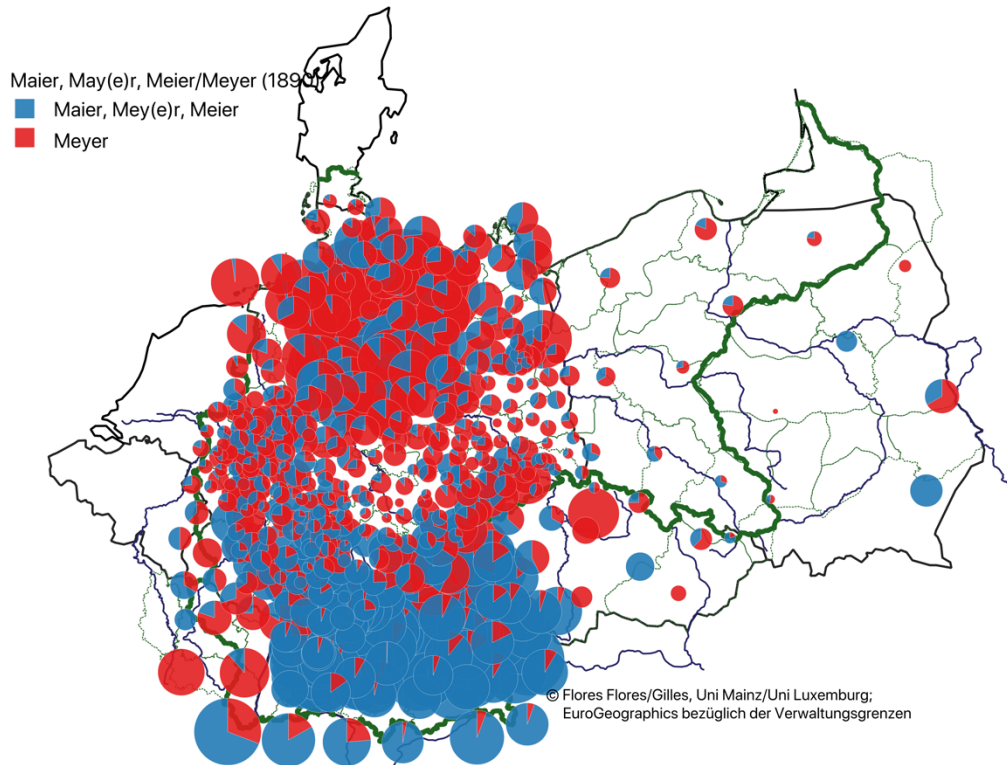


Abb. 5: relative Verbreitung der FamilienN *Maier*, *Mayer*, *Meier*, *Mayr* vs. *Meyer* (Max. Symbolgröße: 33, entspricht 68,01‰) (1890)

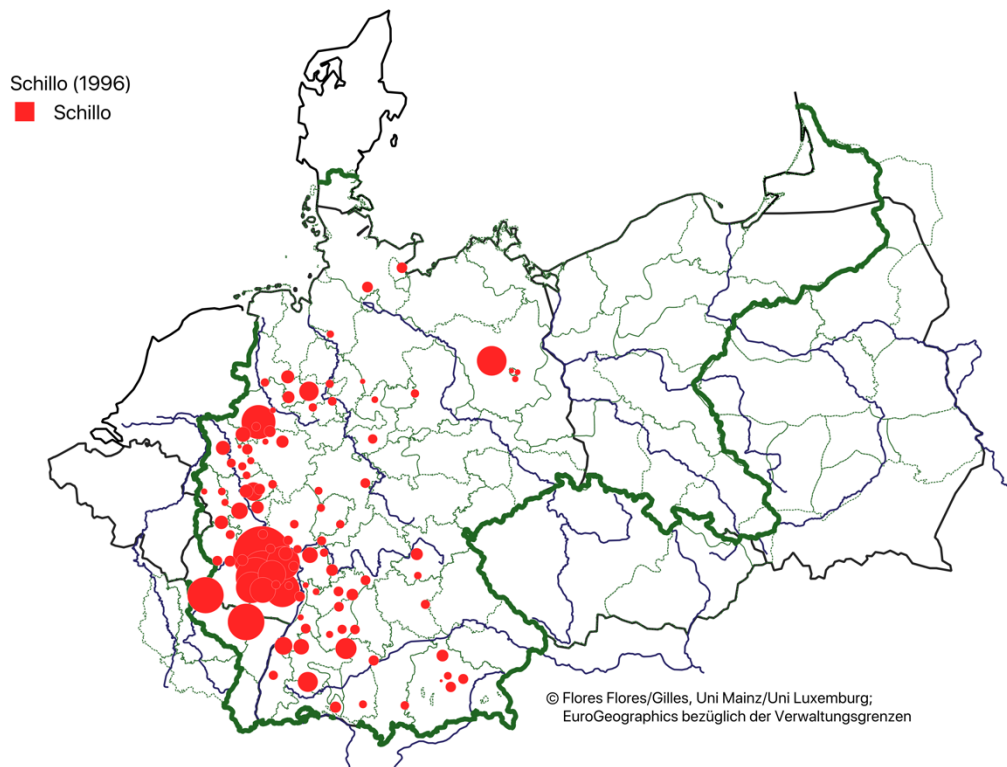


Abb. 6: relative Verbreitung des FamilienN *Schillo* (Max. Symbolgröße: 12, entspricht 25,78‰) (1996)

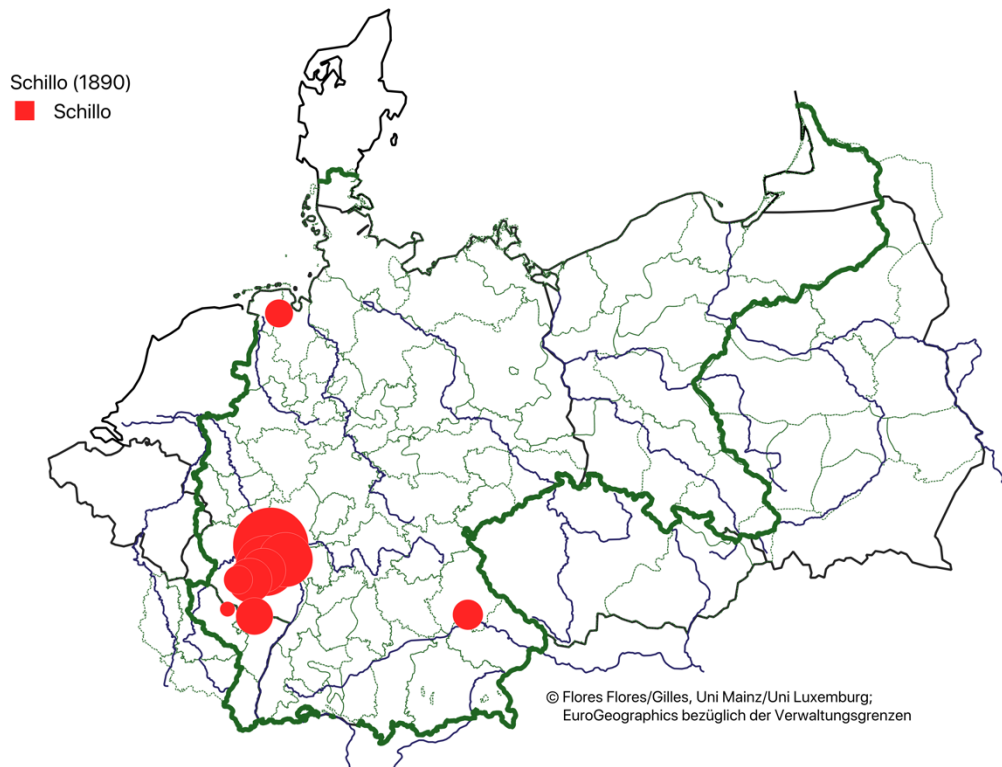


Abb. 7: relative Verbreitung des FamilienN *Schillo* (Max. Symbolgröße: 42, entspricht 90,91%) (1890)

### 3.2. Migrationseinfluss auf die Namenverbreitung

Es lassen sich jedoch auch eine ganze Reihe von FamilienN finden, deren Verbreitung sich 1890 und 1996 (mitunter deutlich) voneinander unterscheidet, d.h. an denen sich jüngere Migrationsbewegungen (v.a. Fluchtbewegungen in Folge des Zweiten Weltkrieges) abzeichnen. Hierunter fallen etwa eine Reihe von slaw. FamilienN, wobei speziell polnische FamilienN bereits vor dem Ersten Weltkrieg in größerer Zahl nach Deutschland, insbesondere ins Ruhrgebiet gelangten:

Nicht erst im 20. Jh., sondern schon Jahrzehnte zuvor, beginnend in der Mitte des 19. Jhs., wanderten Arbeiter aus Polen und den damals preußischen Ostgebieten nach Deutschland ein, vornehmlich in das Ruhrgebiet. [...] Man schätzt ihre Zahl auf ca. 400.000 Menschen. Nicht wenige von ihnen trugen polnische Namen. (Udolph 2011: 184)

Diese frühen poln. Arbeitsmigranten schlugen sich bereits in den Namenverbreitungen von 1890 nieder, da dies sog. „Ruhrpolen nicht die polnische, sondern die deutsche Staatsangehörigkeit mit allen dieselbe ausmachenden Rechten“ (Rymut & Hofmann 2006: VIII) hatten, weshalb sie in den Verlustlisten verzeichnet sind.

Eine zweite große Zuwanderung erfolgte am Ende des Zweiten Weltkrieges und hielt noch in den ersten Nachkriegsjahren an. Mehr als 11 Millionen Deutsche aus Ostpreußen, Pommern, Schlesien, dem Sudetenland und aus anderen Gebieten Ostmitteleuropas wurden als Flüchtlinge oder Vertriebene umgesiedelt. Sie brachten als Deutsche in beachtlichem Umfang auch slawische FN mit (Hubert 1998). (Hengst 2001: 216)

Wie sich beide Migrationswellen auf die slaw. (insbes. poln.) FamilienN in Deutschland auswirken, soll in Abb. 8 und Abb. 9 anhand zahlreichen slaw. FamilienN auf die sehr häufigen slaw. onymischen Suffixe *-ski* bzw. *-sky* veranschaulicht werden, vgl. hierzu auch Nübling, Fahlbusch & Heuser <sup>2</sup>2015: 156–157 (Abb. 26)). Zunächst fällt auf, dass 1890 auch in Tschechien die von Nübling, Fahlbusch & Heuser (<sup>2</sup>2015: 156) als „eher poln.“ bezeichnete Variante *-ski* (zumindest unter der deutschen

Bevölkerung) weitgehend dominiert, „eher tschech.“ *-sky* nur im äußersten Osten – allerdings sind sich beide in den zum Deutschen Reich gehörenden Teilen Polens (in der deutschen Bevölkerung) sehr selten; zu den enorm hohen Frequenzen außerhalb des Deutschen Reichs vgl. Fn. 6. Schon 1890 finden sich zudem – als Folge der ersten Migrationswelle – auch im heutigen Bundesgebiet höhere Konzentrationen v.a. der Variante *-ski*, und zwar wie erwartet v.a. im Ruhrgebiet, daneben auch in den größeren Städten im Norden sowie bei München. In der Fläche sind beide Varianten hingegen selten, kommen aber praktisch überall vor, wobei hier auffälligerweise eher *-sky* dominiert.

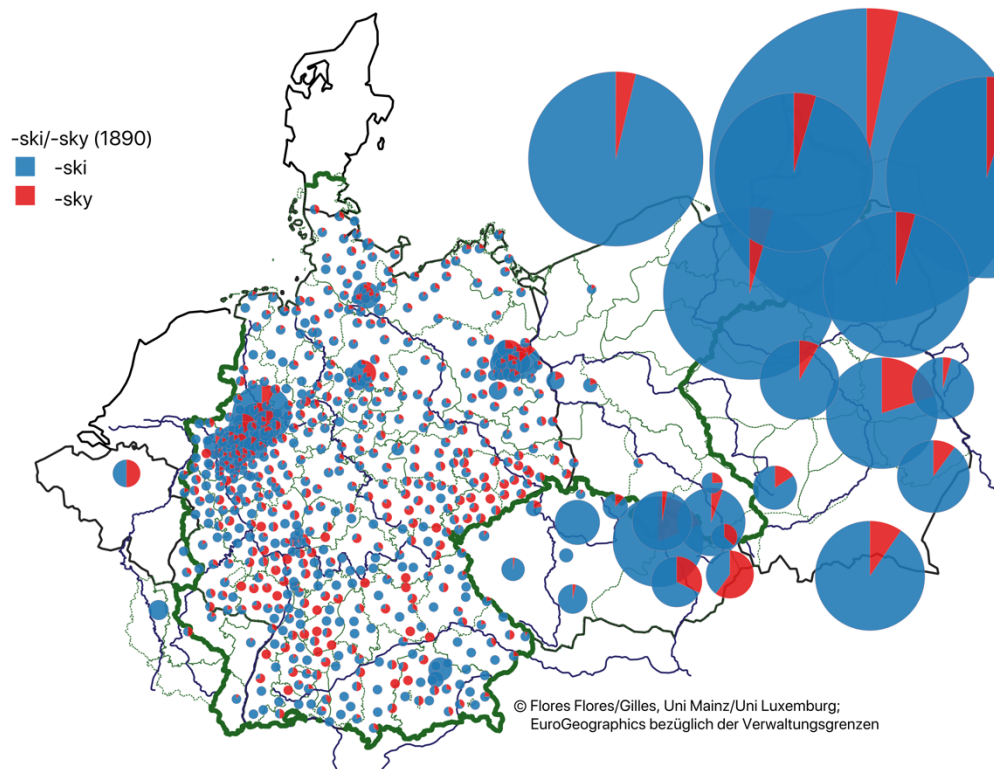


Abb. 8: relative Verbreitung der FamilienN auf *-ski* vs. *-sky* 1890 im Deutschen Reich (Durchmesser der Symbole 2-65, 65 entspricht 338,12‰)<sup>8</sup>

Im Kartenbild von 1996 hingegen zeigt sich auch die zweite Migrationswelle nach dem Zweiten Weltkrieg: die Namen sind nun im gesamten Bundesgebiet insgesamt häufiger, ballen sich aber weiterhin v.a. im Norden.

<sup>8</sup> Hier und in Abb. 9 liegt der Skalierung der Tortensymbole (wie etwa auch beim DFA) der Durchmesser statt wie sonst die Fläche zu Grunde.

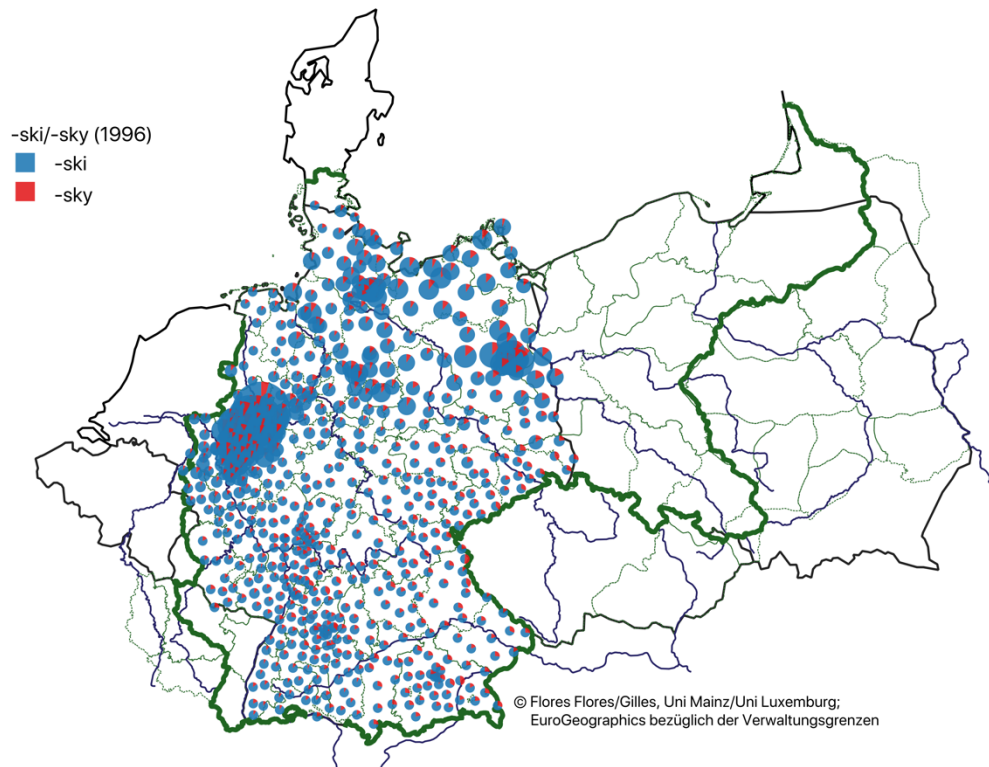


Abb. 9: relative Verbreitung der FamilienN auf *-ski* vs. *-sky* 1996 in Deutschland (Durchmesser der Symbole 2-12, 12 entspricht 62,34%)

Noch deutlicher zeigt sich das Zusammenspiel der beiden Migrationswellen, wenn man einzelne FamilienN in den Blick nimmt: In Abb. 10 wird *Nowak*, ein ÜberN slaw. Ursprungs (zu sorb., poln. *nowak* 'Neuling, Neumann') für einen Neusiedler, daneben auch HerkunftsN zu SiedlungsN mit demselben Ausgangswort im Stamm (vgl. DFA Bd. 2: 769) seiner Variante mit *w*-Ausfall *Noack* für 1996 gegenübergestellt, vgl. zur Genese der Variante und zu zahlreichen weiteren Varianten DFA Bd. 2: 768-773 (K. 348). Im Vergleich mit der Verbreitung beider Namen nach den Verlustlisten (vgl. Abb. 11) zeigt, dass die heute dominante Variante *Nowak* erst in der zweiten Migrationswelle und flächendeckend in die Bundesrepublik gekommen ist, und zwar aus den ursprünglichen Verbreitungsgebieten in Polen und Tschechien. Die einzige Ballung in Deutschland findet sich – als Folge der ersten Migrationswelle – 1890 hingegen im Ruhrgebiet.

Zugleich zeigt sich, dass *Noack* schon 1890 nahezu in gleicher Konzentration und an gleichem Ort, nämlich im sorbisch-sprachigen Raum in Brandenburg und Sachsen, zu finden ist (und dort vor 1890 nahezu exklusiv). Der *w*-Schwund im FamilienN kann somit vor 1890 datiert und als spezifisch sorbisch identifiziert werden. Erwähnenswert ist schließlich, dass die (nicht kartierten) Varianten mit *v* (Typ *Novak*) 1890 im gesamten Kartierungsgebiet inkl. Tschechien praktisch nicht vorkommen, obwohl das entsprechende tschech. Appellativ *novak* lautet.

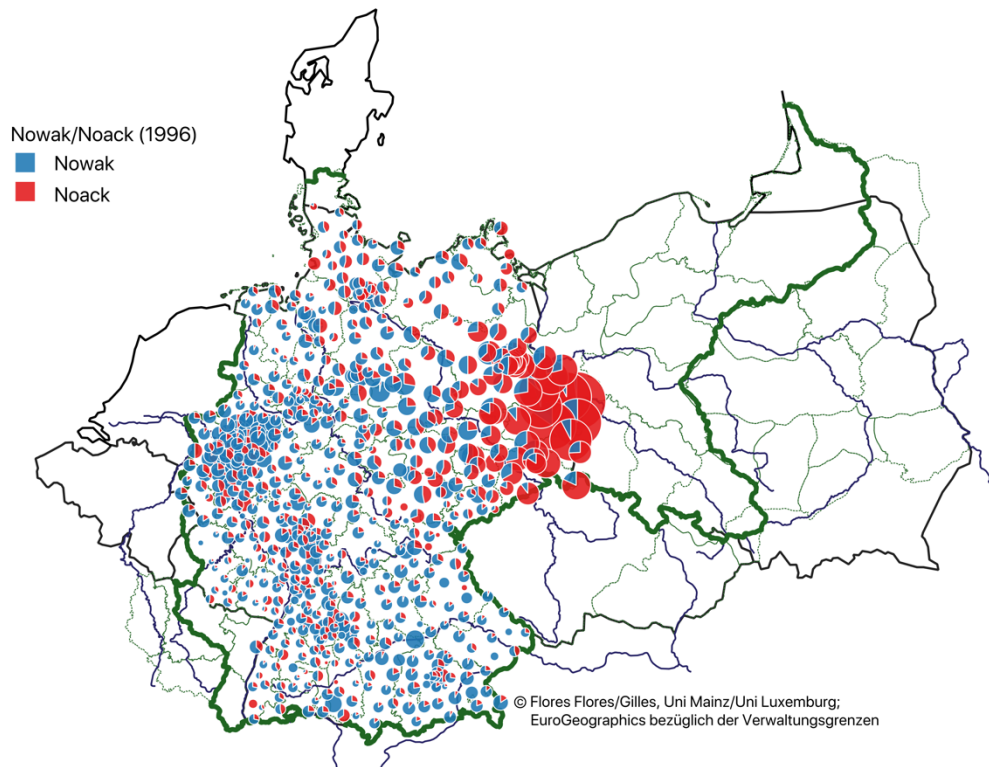


Abb. 10: relative Verbreitung der FamilienN *Nowak* und *Noack* 1996 in Deutschland (Max. Symbolgröße: 12, entspricht 10,41%<sup>9</sup>)

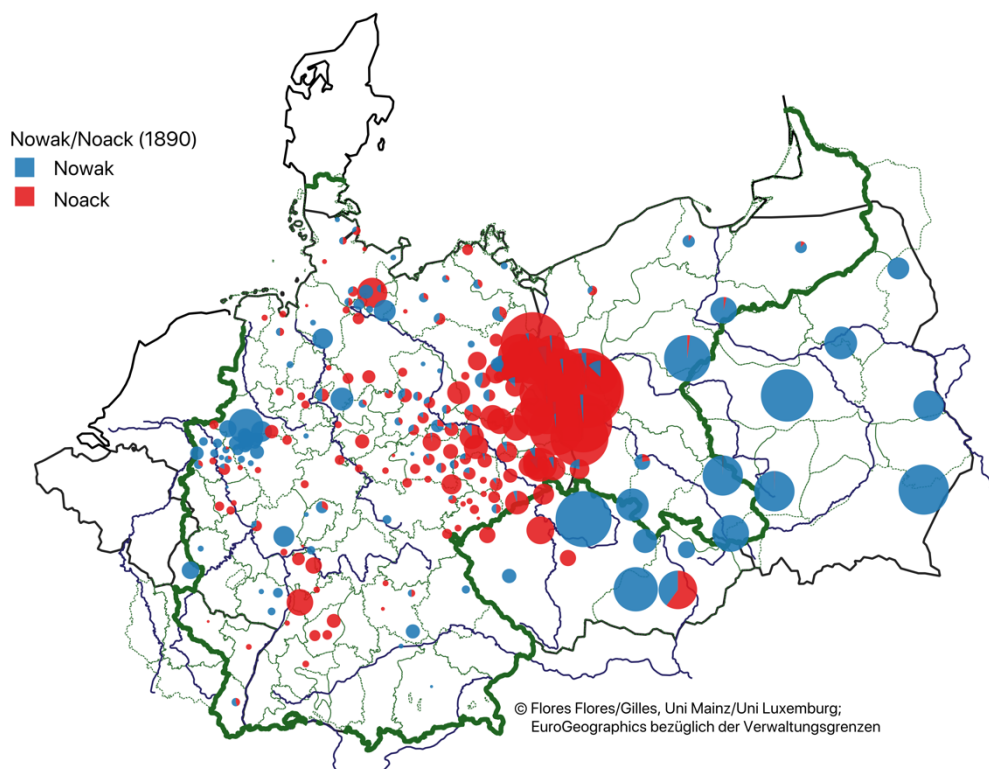


Abb. 11: relative Verbreitung der FamilienN *Nowak* und *Noack* 1890 im Deutschen Reich (Max. Symbolgröße: 22, entspricht 19,48%<sup>9</sup>)

<sup>9</sup> Die Angabe der Symbolgröße bezieht sich auf die Fläche der Kreisdiagramme. Die Maßstäbe der Kreissymbole sind in Karten zu gleichen Namen hier und stets gleich, um eine optimale Vergleichbarkeit zu gewährleisten; d.h. die Größe der Symbole in Abb. 10 und Abb. 11 sind bei gleicher relativer Häufigkeit gleich.



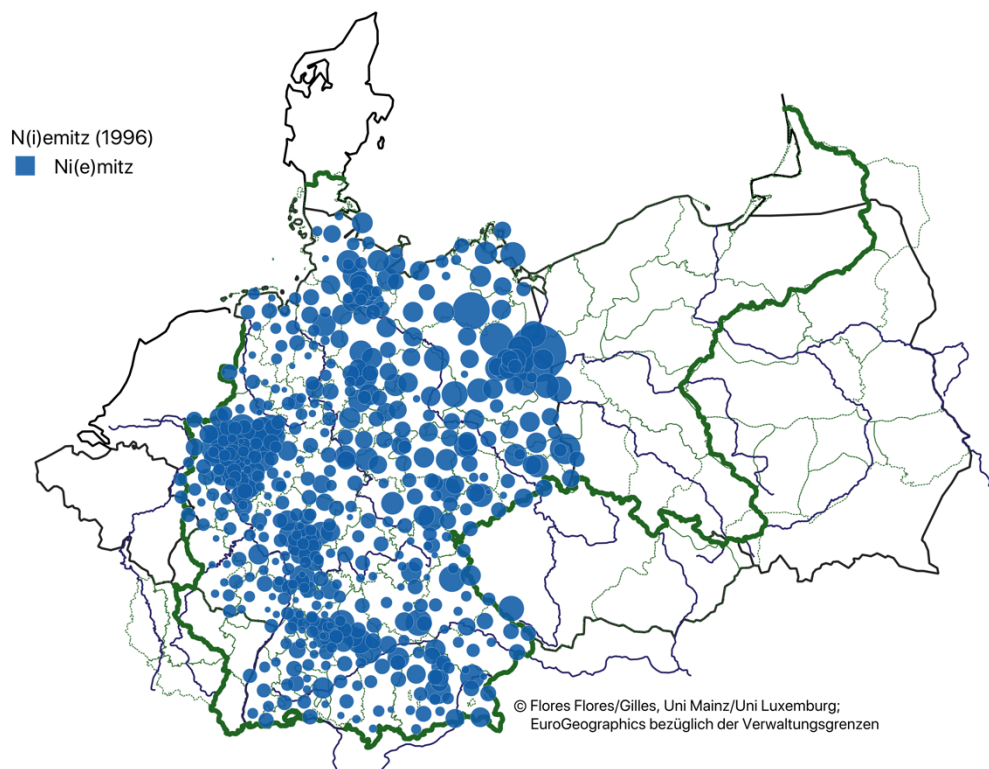


Abb. 12: relative Verbreitung des FamilienN-Typs *N(i)emitz* 1996 in Deutschland (Max. Symbolgröße: 12, entspricht 1,35‰)

Anders hingegen die FamilienN des Typs *N(i)emitz*, die auf poln. *Niemiec*, obersorb. *Němc* [iə], niedersorb. *Nimc* und tschech. *Němec* 'Deutscher' zurückgehen (vgl. DFA Bd. 4, 9–11), Bedeutungskonkurrenzen bestehen v.a. mit gleichlautenden SiedlungsN in Sachsen und Schlesien (vgl. DFA Bd. 4, 9–10). In Abb. 12 sind die Varianten *Niemi(e)tz*, *Niemi(e)tz*, *Niemetz*, *Niemi(e)c*, *Niemec*, *Ni(e)m(t)z*, *Ni(e)m(t)sch(ke)* sowie *Nem(h)itz*, *Neme(t)z*, *Nemec(z)(ek)*, *Nemetschek*, *Nemetschke*, *Nemesch* für 1996 kartiert. Die FamilieN sind heute in Deutschland ubiquitär, mit Ballungen in den Metropolregionen v.a. an Ruhr und Rhein sowie im äußersten Osten.

Im Vergleich mit der Verbreitung von 1890 (vgl. Abb. 13) zeigt sich, dass die entsprechenden FamilienN im Wesentlichen erst in der zweiten Migrationswelle nach Deutschland gekommen sind: Außer in Grenznähe zum heutigen Polen, wo sie wohl überwiegend dem Sorbischen zuzuordnen, d.h. nicht migriert sind, und östlich von Hannover kommen sie noch kaum vor. Deutlich zeigen sich 1809 die arealen Unterschiede zwischen den *e*- und *i*-haltigen Typen *Nemitz* bzw. *Ni(e)mitz* (zur heutigen Verbreitungsdifferenz vgl. DFA Bd. 4: 9–13 (K. 5 und 6)): Während im heutigen Polen (vgl. poln. *Niemiec*) Typ *Ni(e)mitz* dominiert (mit Ausnahme der Woiwodschaft Pommern, dort mehrheitlich Typ *Nemitz*), findet sich im heutigen Tschechien fast ausnahmslos Typ *Nemitz* (vgl. tschech. *Němec*). Erstaunlich ist dagegen der Befund im heutigen Deutschland, wo beide Typen v.a. im sorbischen Gebiet vorkommen: Im Bereich des obersorb. kommt passend zu obersorb. *Němc* [ie] praktisch ausschließlich Typ *Niemitz* vor, im niedersorb. Areal (vgl. niedersorb. *Nimc*) hingegen dominiert stellenweise Typ *Nemitz*.

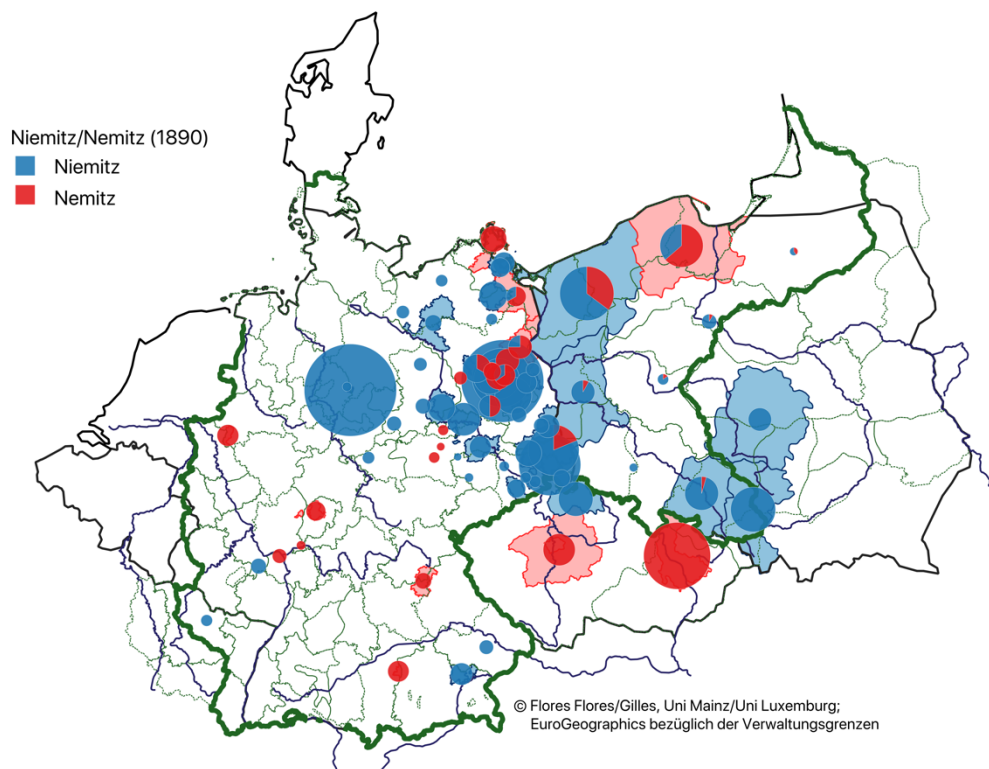


Abb. 13: relative Verbreitung der FamilienN-Typen *Ni(e)mitz* vs. *Nemitz* 1890 im Deutschen Reich (Max. Symbolgröße: 30, entspricht 3,41‰; Anzeigeschwelle der Flächen: 0,1‰)

Zusätzlich zu den Kreis- bzw. Tortendiagrammen ist in Abb. 13 der jeweils dominante Typ durch eine Einfärbung in der Fläche markiert, die Anzeigeschwelle liegt bei 0,1‰, d.h. dass eine Einfärbung nur erfolgt, wenn mindestens die dominante Variante eine relative Häufigkeit von mehr als 0,1‰ aufweist. Eine weitere Möglichkeit der Visualisierung zeigt Abb. 14, in der die Daten von 1996 und 1890 in einer Karte kombiniert werden: Abgebildet wird die Veränderung der relativen Häufigkeit von 1890 gegenüber 1996, wobei hier wiederum zwischen den beiden Untertypen *Niemitz* und *Nemitz* nicht unterschieden wird. Zuwachs und Absinken der Frequenzen werden jeweils in vier Abstufungen durch unterschiedlich intensive Einfärbung der Fläche symbolisiert. Es wird ersichtlich, dass Typ *N(i)emitz* fast im gesamten Bundesgebiet 1996 gegenüber 1890 deutlich häufiger geworden ist, gesunken sind die Frequenzen lediglich im Bereich der ursprünglichen Nester, am stärksten in Hannover.

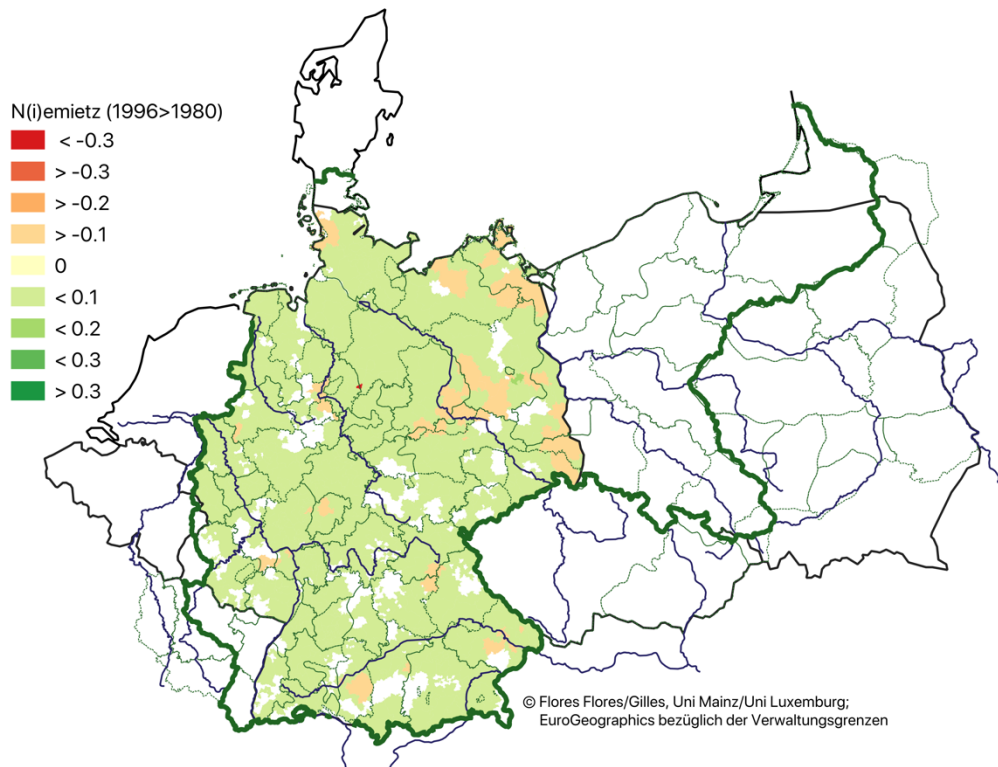


Abb. 14: Veränderung der relativen Häufigkeit des FamilienN-Typs *N(i)emitz* 1890–1996

Abb. 15<sup>10</sup> und Abb. 16 zeigen die Verbreitung der beiden FamilienN *Böhm* und *Böhme* 1996 und 1890. Es handelt sich dabei schon 1890 wie heute um die „beiden mit Abstand vorherrschenden Varianten“ (DFA Bd. 4: 87) der HerkunftsN zu *Böhmen*, wobei die einsilbige Variante direkt auf ahd. *Bēheim* 'Böhme, der aus Böhmen' zurückzuführen ist, und *Böhme* daraus durch sekundären *e*-Antritt entstanden ist (vgl. im Detail sowie zur Genese des Vokalismus *ö* Kunze 2001: 424 und DFA Bd. 4: 87; eine Kartierung der selteneren Varianten dieses HerkunftsN findet sich in DFA Bd. 4: 90 (K. 36), Einzelkarten aller Varianten bei Kunze 2001).

<sup>10</sup> Vgl. DFA Bd. 4: 86 (K. 35).

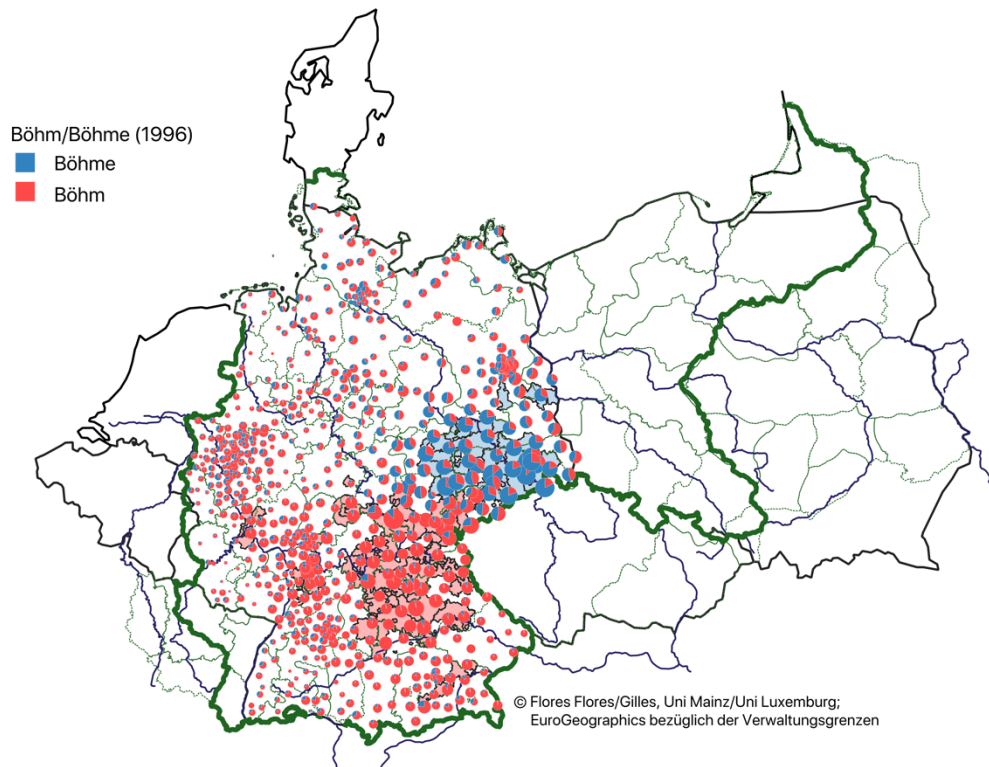


Abb. 15: relative Verbreitung der FamilienN *Böhm* vs. *Böhme* 1996 in Deutschland (Max. Symbolgröße: 3, entspricht 5,32‰; Anzeigeschwelle der Flächen: 1,5‰)

Die Variante *Böhm* dominiert insgesamt 1996 wie 1890. Die Verbreitung von *Böhme* mit einem Kern in Sachsen und im südlichen Brandenburg und Sachsen-Anhalt bleibt weitgehend stabil und entspricht auch den Erstbelegen „seit Mitte des 14. Jh. im ostmd. Raum und Ostfalen“ (DFA Bd. 4: 87), wobei 1890 das Kernverbreitungsgebiet bis ins nordwestliche Böhmen reicht.<sup>11</sup> Die Verbreitung von *Böhme* hingegen unterscheidet sich 1890 und 1996 stärker: 1890 ist auch diese Variante noch auf ein klareres Gebiet beschränkt, dessen Kern sich einerseits südlich ans Verbreitungsgebiet von *Böhm* anschließt und von dort nach Westen bis ins Rhein-Main-Gebiet zieht, andererseits im heutigen Tschechien und Polen (nord-)östlich an Böhmen anschließt. Beide Varianten legen sich also 1890 in einem Ring um Böhmen, wobei *Böhm* zwischen den beiden Zentren von *Böhme* liegt. Die heute neben dem ursprünglichen Kern über das gesamte verbreiteten *Böhme*-Belege lassen sich mutmaßlich auf Nachkriegsmigration der entsprechenden Namenträger im heutigen Tschechien und Polen zurückführen.

<sup>11</sup> Dies könnte ein Beleg für Walthers (1993) These sein, das *Unger/Hunger* und *Böhm(e)* FamilienN sächsischer Bergleute sind, die in ungar./böhm. Bergbauggebiete abgewandert und zurückgekehrt sind, vgl. DFA Bd. 4: 37.

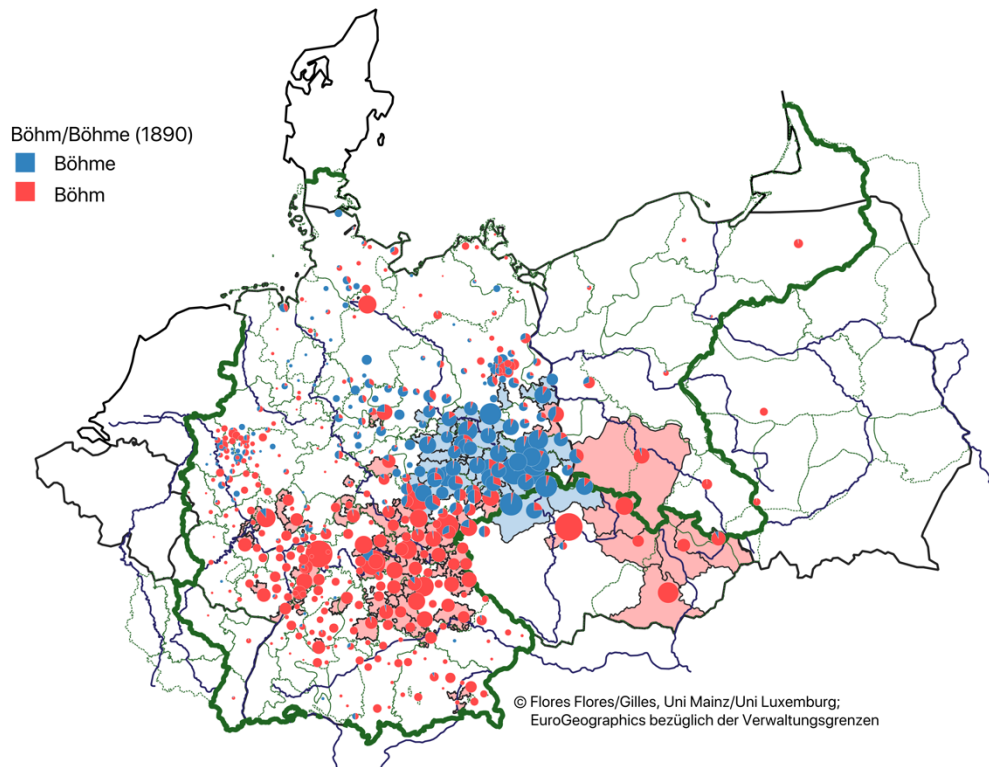


Abb. 16: relative Verbreitung der FamilienN *Böhm* vs. *Böhme* 1890 im Deutschen Reich (Max. Symbolgröße: 6, entspricht 10,40‰; Anzeigeschwelle der Flächen: 1,5‰)

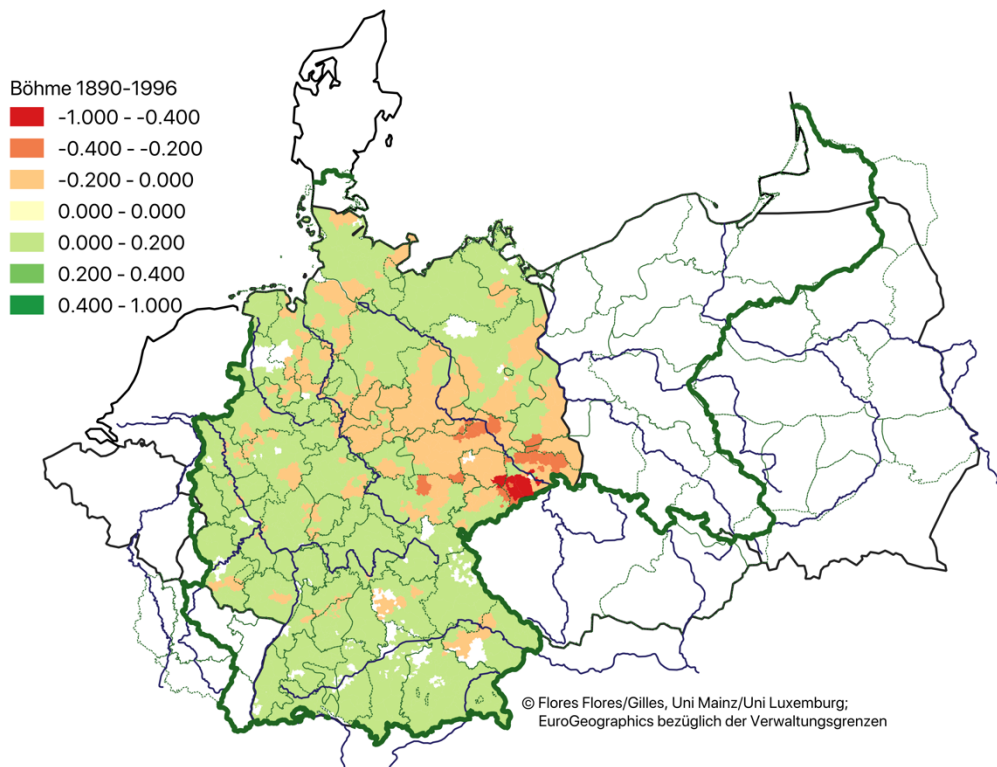


Abb. 17: Veränderung der relativen Häufigkeit des FamilienN *Böhme* 1890–1996

Dass auch *Böhme* sich 1996 in Deutschland stärker in die Fläche ausgebreitet hat, während zugleich die höchsten Konzentrationen im alten Kerngebiet gesunken sind, wird in der Kombinationskarte durch die

höheren Belegzahlen von *Böhm* überlagert, zeigt sich aber wiederum sehr klar in einer Veränderungskarte (nur Variante *Böhme*), vgl. Abb. 17.

Ein neues Licht werfen die historischen Daten auch auf die Verbreitung der Apokope in den beiden FamilienN-Varianten *Schulz* und *Schulze* (BerufsN aus ahd. *sculdheizo*, as. *Sculdhetio*; im Detail vgl. Kunze & Kunze 2003: 135; dort auch eine entsprechende Verbreitungskarte (Kunze & Kunze 2003: 188 (K.9)), und zwar sowohl in ihrem Kernverbreitungsgebiet, das Kunze & Kunze (2003: 169) im Ostmitteldeutschen verorten, als auch in der Fläche mit geringerer Frequenz, vgl. Abb. 18 und Abb. 19. In beiden Karten sind die Flächen gestuft eingefärbt: Schwache Färbung bedeutet, dass die betreffende Namensvariante dort mit einer höheren Frequenz vorkommt, als die andere. Durch dunkle Einfärbung wird zusätzlich hervorgehoben, wo die dominante Variante mit einer relativen Häufigkeit von über 1,0 % auftritt, d.h. wie die jeweiligen Verbreitungsschwerpunkte liegen. Es zeigt sich, dass für beide Varianten zusammengenommen die Verortung des Hauptverbreitungsgebiets durch Kunze & Kunze (2003) schon 1890 bestätigt werden kann. Allerdings fällt auf, dass 1890 sowohl das in Abb. 19 dunkel hervorgehobene Kernverbreitungsgebiet Variante *Schulze*, als auch das schwach eingefärbte Gebiet, in dem diese unapokopierte Variante bei geringen Gesamtfrequenzen dominiert, in den heutigen Grenzen Deutschlands jeweils deutlich größer sind. Das Hauptverbreitungsgebiet der heute insgesamt dominanten apokopierten Variante *Schulz* hingegen ist 1890 im Osten in den heutigen poln. Woiwodschaften Westpommern, Pommern und Lebus fortgesetzt; nicht-apokopierte Formen kommen in den ehemaligen Ostgebieten 1890 hingegen praktisch nicht vor.

Die heutige Dominanz der apokopierten Form wird auch hier mutmaßlich wieder zumindest auch der Nachkriegsmigration aus dem östlichen Verbreitungsgebiet von 1890 geschuldet sein. U.U. findet sich hier auch eine Antwort auf die von Kunze/Kunze (2003: 169) gestellte Frage finden, „warum [...] trotz dieser [Benennungsmotiv, phonotaktischem Aufbau, Hauptverbreitungsgebiet, vorhandene Varianten] gleichartigen Bedingungen der Anteil apokopierter Varianten im betreffenden PLZ 0 bei *Schulz*, *Scholz* mit 59,4 % doppelt so hoch wie bei *Schütz*, mit 29,2 % [...]“ ist. Ein Abgleich der historischen Verbreitungen weiterer ähnlich gelagerter FamilienN kann sich hier als aufschlussreich erweisen.

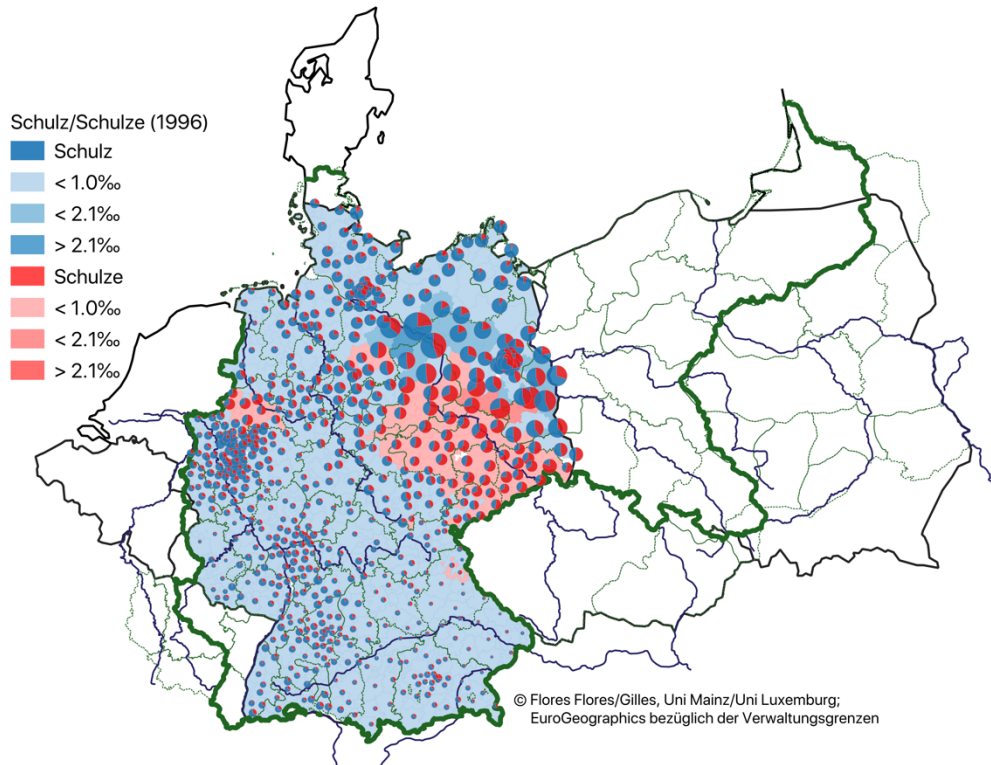


Abb. 18: relative Verbreitung der FamilienN *Schulz* vs. *Schulze* 1996 (Max. Symbolgröße: 6, entspricht 37,74 Flächen gestaffelt)

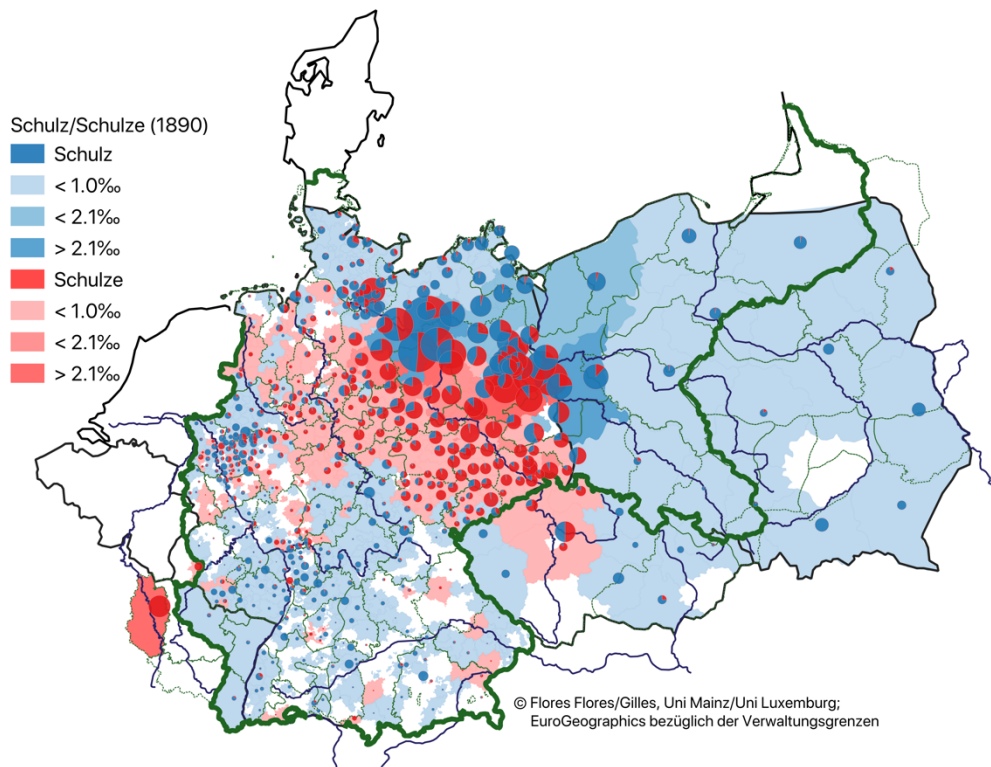


Abb. 19: relative Verbreitung der FamilienN *Schulz* vs. *Schulze* 1890 im Deutschen Reich (Max. Symbolgröße: 17, entspricht 100,57‰; Flächen gestaffelt)

## 4. Isonymieanalyse

Abschließend soll mit der sog. Isonymieanalyse eine weitere Methode zur Analyse georeferenzierter Namendaten vorgestellt werden und auf die vorgestellten Namendaten angewandt werden. Es handelt sich um eine quantitative Methode, mit der größere Namen-Datensätze ohne Vorannahmen, d.h. explorativ, auf räumliche Strukturen untersucht werden können. Auch wenn sie in der dt.-sprachigen Onomastik bislang kaum Verwendung gefunden hat, deuten sich die Möglichkeiten der Isonymieanalyse für die FamilienN-Geographie in etlichen internationalen Arbeiten, denen allerdings weitgehend eine linguistisch-onomastische Perspektive fehlt, insofern sie sich mit dem Aufzeigen (tlw. nicht überraschender) räumlicher Strukturen begnügen, ohne nach den dahinterstehenden Phänomenen zu fragen.<sup>12</sup> Diese Lücke schließen kann die von Flores Flores (i. Vorb./2015) sog. Häufigkeitsdifferenzanalyse, mit der die maßgeblich für die Raumbildung verantwortlichen FamilienN ermittelt werden können. Im Folgenden werden die grundlegenden Konzepte hinter der Isonymieanalyse erläutert und mit Hilfe der Namendaten veranschaulicht. Auch hier können erste Ergebnisse nur angedeutet werden und sollen in erster Linie Anregung für weitere Arbeiten sein.

### 4.1. Laskerdistanz $L_{ij}$

Basis der Isonymieanalyse ist die von Lasker (1985) entwickelte Laskerdistanz  $L_{ij}$ , mit deren Hilfe die (Un)Ähnlichkeit beliebiger Nameninventare anhand der gemeinsamen Namen ermittelt wird.<sup>13</sup> Die Berechnung basiert dabei auf den Gesamtinventaren, d.h. auf allen jeweils vorkommenden FamilienN, und nicht lediglich auf einer geringen Stichprobe wie etwa beim Vergleich der häufigsten Namen oder auch randomisierter Stichproben,<sup>14</sup> berücksichtigt wird zudem nicht nur das Vorhandensein gemeinsamer Namen in mehreren Teilinventaren, sondern die jeweilige Frequenz der Namen in ihnen: Je größer die Tokenzahl eines Namen in zwei Inventaren, umso stärker beeinflusst dieser Name deren Ähnlichkeit im Sinne der Laskerdistanz positiv.<sup>15</sup>

Konkret wurden die Namendaten im Bereich der heutigen Bundesrepublik nach 2-stelligen PLZ in 95 Teilgebiete/-inventare untergliedert; zu diesen kommen für die Daten der Verlustlisten noch neun Teilgebiete/-inventare im heutigen Polen, drei im heutigen Frankreich sowie je eines im heutigen Belgien, Luxemburg, Tschechien sowie in den heutigen Niederlanden (=insgesamt 111 Teilgebiete/-

---

<sup>12</sup> CHESHIRE/LONGLEY/MATEOS (2010), CHESHIRE/MATEOS/LANGLEY (2009), SCAPOLI ET AL. (2007); VIDESOTT (2014), der einen Überblick über „onomatometrische“, d.h. statistisch-mathematische Methoden der Onomastik gibt. Zwar verweist er in den Fn. 5 und 6 auf einige Arbeiten, vgl. Flores Flores (i. Vorb./2015: 13, ??).

<sup>13</sup> ISONYMIE-BEGRIFF; Humangenetik; Unsauberkeiten, hier aber unwichtig

<sup>14</sup> ÜBLICHE VERDÄCHTIGE?

<sup>15</sup> Berechnet wird die Laskerdistanz immer paarweise für zwei Teilinventare nach der Formel

$$L_{ij} = -\ln\left(2 \times \frac{\Sigma(S_i \times S_j)}{2 \times \Sigma S_i \times \Sigma S_j}\right)$$

mit  $i$  und  $j$  als zwei von beliebig vielen Ortspunkten/Teilinventaren, in die das zu untersuchende Nameninventar untergliedert wird, und  $S$  als Anzahl der Namenträger für jeden einzelnen Namen in diesen beiden Ortspunkten  $i$  und  $j$ . Näheres zum Ursprung des Konzepts sowie zur Berechnung und Implementierung der Laskerdistanz vgl. Flores Flores (i. Vorb./2015 55–57).



inventare). Zwischen all diesen muss dann kreuzweise die Laskerdistanz ermittelt werden, was insgesamt über 6100 Einzel-Laskerdistanzen (in einer 111-dimensionalen Matrix) ergibt.

## 4.2. Clusteranalyse nach *Ward*

Die weitere Verarbeitung der Laskerdistanzen erfolgt mit Hilfe des an der Universität Groningen entwickelten Online-Analysetool Gabmap, mit dem neben Dialektdaten auch Differenzdaten (wie die Laskerdistanzen) analysiert und kartographisch dargestellt werden können.<sup>16</sup> Von den vielfältigen statistisch-mathematischen Modellen, die für die Analyse zur Verfügung stehen, soll hier stellvertretend die Clusteranalyse mit dem *Ward*-Algorithmus vorgestellt werden.<sup>17</sup>

Beim *Ward*-Algorithmus handelt es sich um eines von zahlreichen Modellen zur hierarchischen Cluster-Analyse. Bei dieser werden die einzelnen Teilinventare nach ihrer Ähnlichkeit in verschiedenen Clustern gruppiert. Die Gruppierung erfolgt Schrittweise, beginnend mit den beiden Teilinventaren mit der geringsten Laskerdistanz; im Wesentlichen unterscheiden sich die verschiedenen Algorithmen darin, wie im Folgenden die Distanz zwischen einzelnen Teilinventaren und bereits gruppierten Clustern (bzw. zwischen mehreren gruppierten Clustern) berechnet wird.<sup>18</sup> Der *Ward*-Algorithmus<sup>19</sup> zeichnet sich dabei dadurch aus, dass er neben den geometrischen Distanzen auch die Größe der Cluster berücksichtigt, was zu verhältnismäßig gleich großen Clustern führt.<sup>20</sup>

Die Zahl der Cluster, in die der Datensatz aufgeteilt wird, wird nicht vom Algorithmus bestimmt, sondern kann prinzipiell beliebig festgelegt werden. Wichtig ist, nicht zu viele Cluster zu wählen, da sonst entweder eigentlich sehr ähnliche Teilinventare in verschiedene Cluster getrennt werden, oder aber im schlimmsten Fall sogar einzelne Teilinventare anderen in einem anderen Cluster ähnlicher sind als denen im eigenen Cluster. Um die maximale Anzahl der sinnvoll unterscheidbaren Cluster zu ermitteln, sollte einerseits die kophänetische Distanz der einzelnen Cluster und Teilgebiete beachtet werden, die im Wesentlichen anzeigt, wie ähnlich/unähnlich sich Cluster/Teilinventare (nach dem zugrundeliegenden Algorithmus) sind (vgl. Flores Flores 2015/i Vorb.: 77, Cheshire, Mateos & Longley 2009: 23 und Kleiweg, Nerbonne & Bosveld 2004). Andererseits kann auch mit Hilfe einer sogenannten Multidimensionalen Skalierung (MDS) überprüft werden, ob die Gruppierungen der Clusteranalyse in den Ausgangsdaten tatsächlich angelegt sind. Es handelt sich hierbei um ein mathematisches Verfahren, um komplexe, mehrdimensionale Verteilungsdaten (wie die Laskerdistanzen) mit möglichst geringem

---

<sup>16</sup> GABMAP; für eine ausführliche technische Beschreibung und der Anwendungsmöglichkeiten von Gabmap vgl. Nerbonne et al. (2011).

<sup>17</sup> Generell „gibt es keine ‚richtigen‘ oder ‚falschen‘ Klassifikationsverfahren (und darauf beruhende Klassifikationsergebnisse), sondern nur solche, die ‚brauchbarer‘, ‚zweckdienlicher‘ oder ‚optimaler‘ als andere sind“ (Visedott 2014: 217), so dass die Auswahl ausgehend vom verfolgten Erkenntnisgewinn erfolgen muss. Eine solche teleologische Methodenwahl ist nicht unproblematisch, da unterschiedliche mathematische Modelle wie auch geringfügige Veränderungen einzelner Faktoren die Ergebnisse teilweise deutlich beeinflussen können. Sinnvoll ist es daher, einerseits verschiedene Modelle anzuwenden und deren Ergebnisse abzugleichen, und andererseits, die Korrelation der Ergebnisse mit den Ausgangsdaten mit statistischen Methoden. Beides ist mit Gabmap möglich und hier erfolgt, kann aber allenfalls in Grundzügen erläutert werden.

<sup>18</sup> Vgl. im Detail Flores Flores (i. Vorb./2015 75–77. ->

<sup>19</sup> FORMEL (Flores Flores (i. Vorb./2015 76)

<sup>20</sup> Vgl. etwa Gries 2009: 317.

Informationsverlust auf beliebig wenige Dimensionen zu reduzieren.<sup>21</sup> Auf diese Weise können in einer zweidimensionalen Skalierung (2DS) alle Teilgebiete in einem Punktdiagramm abgebildet werden, in dem ihre Abstände näherungsweise den Laskerdistanzen zwischen ihnen entsprechen. Die nötigen Abweichungen von den Ausgangsdaten kann dabei u.a. im Korrelationskoeffizient  $r$  wiedergegeben werden. Für die Clusterbildung gilt: Wenn sich die Cluster auch in einer entsprechenden 2DS mit möglichst hoher Korrelation als diskrete Bereiche abbilden, sind sie in den Grunddaten angelegt.

Für die Verlustlisten lassen sich so mit dem *Ward*-Algorithmus maximal zehn Cluster sinnvoll unterscheiden. In Abb. 20 sind sie durch verschiedenfarbige Flächen markiert und mit einer groben Gliederung des deutschen Dialektraumes überlagert. Es zeigt sich, dass die Clustergrenzen teilweise gut mit Dialektgrenzen übereinstimmen; gut bilden sich ab die Grenzen zwischen Niederdt. und Hochdt., südlichem West- und Ostniederdt., Niederrhein. und Westfäl. sowie Ostfrk. und Alem./Bair. Die Ost-West-Gliederung des mitteldt. Sprachraums hingegen findet in den Clustern keine Entsprechung, stattdessen bilden sich eher der Nieder-/Mittelrhein und die Mosel ab.

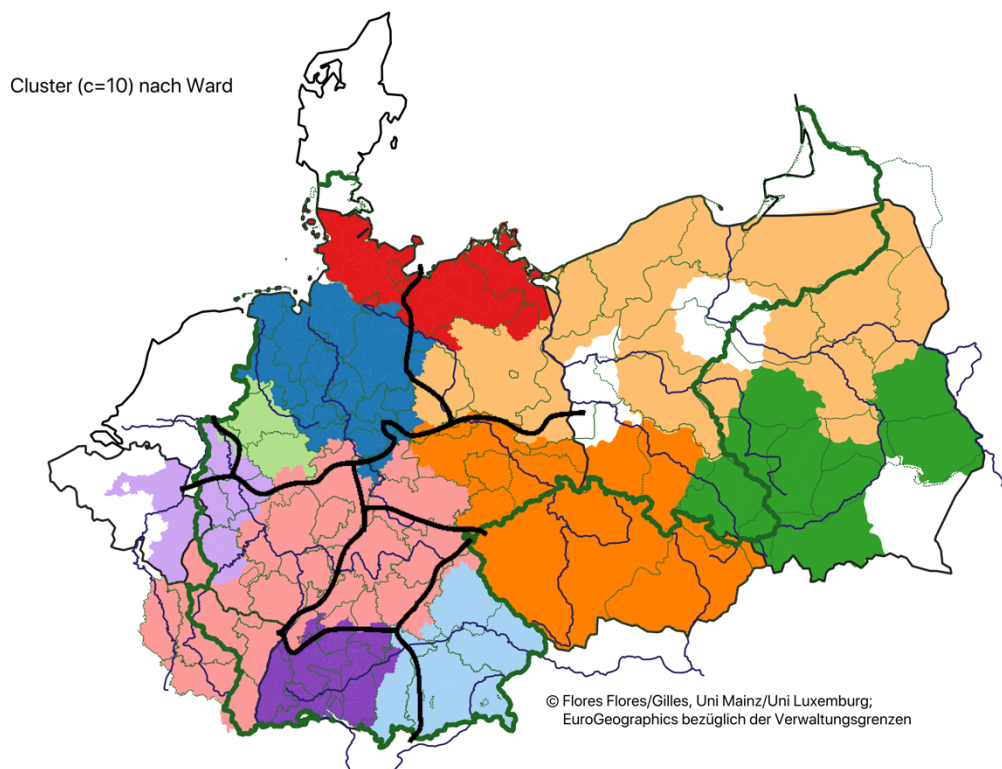


Abb. 20: Clusteranalyse nach *Ward* (C=10 Cluster) mit Einteilung der deutschen Dialektgebiete (1890)

Auch hier soll wieder ein vergleichender Blick auf die rezenten Telefonbuchdaten geworfen werden: Abb. 21 zeigt die Unterteilung der Telefonbuchdaten von 2012 in zwölf Cluster nach *Ward* (nach Flores Flores/Gilles 2014). Auch hier zeigt sich die u.a. Grenze zwischen hoch- und niederdt. Sprachraum,

<sup>21</sup> Dies ist zu vergleichen mit der Übertragung der (dreidimensionalen) geographischen Abstände von der Erdkugel in ein zweidimensionales Kartenmodell: Es ist nicht möglich, alle Städte der Erde auf der zweidimensionalen Projektion gleichzeitig mit dem korrekten Abstand zueinander abzubilden. Näheres zur MDS findet sich u.a. bei Flores Flores (i. Vorb./2015: 65–69 sowie speziell auch zu den Möglichkeiten der Visualisierungen der MDS bei Nerbonne/Heeringa/Kleiweg (1999) und Leinonen (o.J.: 9 und 2010: 207–208).

daneben fallen aber auch einige Unterschiede auf: Das Ostmitteldt. wie das Ostniederdt. bilden jeweils recht exakt einen eigenen Cluster, ebenso fällt die alem.-bair. Grenze mit einer Clustergrenze zusammen. Am auffälligsten sind aber wohl der graue und dunkelgrüne Cluster, die sich als (untypisch für eine *Ward*-Gliederung) als unzusammenhängender „Flickenteppich“ durch den niederdt. Und mestmitteldt. Raum ziehen. Ob/inwiefern sich diese Unterschiede auf Nachkriegs-Migration zurückführen lassen, oder ob sich hier u.U. auch eine dialektale Diachronie abbildet, ist bislang nicht untersucht. Im Vergleich wird jedoch deutlich, dass die Clusteranalyse der Verlustlisten ein deutlich homogeneres Bild abgibt, d.h. dass diese FamilienN-Strukturen die historischen Dialektraumgliederungen konsistenter abbilden.

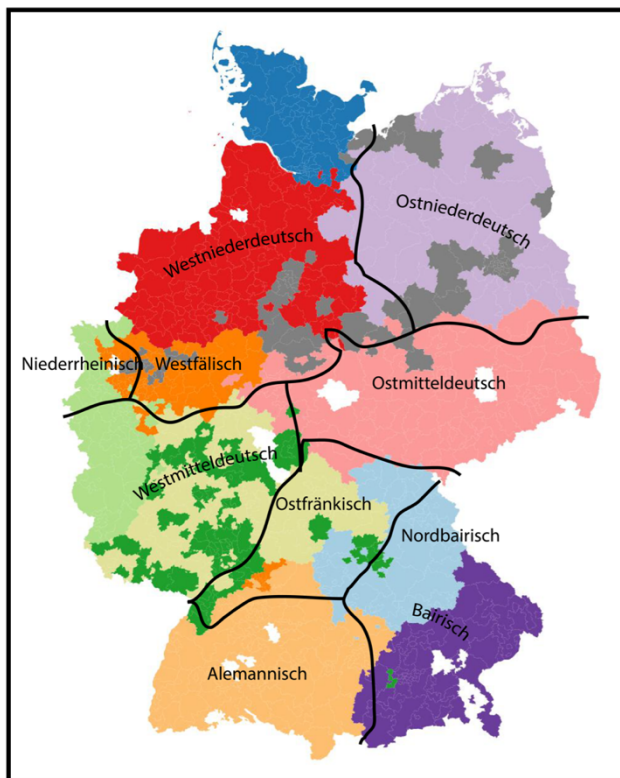


Abb. 21: Clusteranalyse nach *Ward* (C=12 Cluster) mit Einteilung der deutschen Dialektgebiete (2012, nach Flores Flores /Gilles 2014)

### 4.3. Häufigkeitsdifferenzindex (HDI)

In einem letzten Schritt können die rein quantitativen Ergebnisse der Isonymieanalyse um eine qualitative Ebene ergänzt werden, indem die für die Cluster- d.h. Arealbildung relevanten Namen (und damit Phänomene) ermittelt werden. Hierfür kommt das von Flores Flores (i. Vorb./2015: 111–115) entwickelte Konzept des Häufigkeitsdifferenzindex (HDI) zum Einsatz. Er basiert auf der Annahme, dass die Laskerdistanz letztlich auf den Frequenzunterschieden der FamilienN in den untersuchten Gebieten beruht;

vereinfacht gilt: Zwei Nameninventare sind einander dann desto ähnlicher, umso mehr Familiennamen in ihnen beiden häufig vorkommen (d.h. umso [...] kleiner [ist] damit ihre Laskerdistanz [...]). Umgekehrt kann daraus abgeleitet werden, dass in der Isonymieanalyse gerade zwischen zwei solchen Ortspunkten am wahrscheinlichsten Grenzen gebildet werden, bei denen verhältnismäßig viele Familiennamen eine möglichst große Differenz ihrer relativen Häufigkeiten aufweisen. (Flores Flores i. Vorb./2015: 111)

Um also die für die Grenzziehung zwischen zwei Clustern relevanten Namen zu ermitteln, wird zunächst für alle Namen die Differenz der relativen Häufigkeiten in den beiden Clustern berechnet, die sog. Häufigkeitsdifferenz  $\Delta_H$ :

$$\Delta_H(n_{ij}) = \left| \frac{\text{Tokens}(\text{Name}_n)_i}{\text{Namenträger}_i} - \frac{\text{Tokens}(\text{Name}_n)_j}{\text{Namenträger}_j} \right|$$

Anschließend werden die ermittelten Werte durch die durchschnittliche Häufigkeitsdifferenz der beiden Cluster geteilt, um zwischen allen Clustergrenzen vergleichbare Werte zu erhalten, die die Zahl der jeweiligen Namenträger und Häufigkeitsverteilungen berücksichtigt, den sog. Häufigkeitsdifferenzindex HDI:

$$\text{HDI}_n = \frac{\Delta_H(n_{ij})}{\bar{\Delta}_H(ij)}$$

Bezogen auf die Namendaten der Verlustlisten ergibt sich so folgendes Bild: Zunächst wird die Zahl der Cluster der *Ward*-Gliederung auf  $C=3$  reduziert, was er in den Daten am stärksten angelegten Raumstruktur mit der höchsten Korrelation zu den Ausgangsdaten entspricht, vgl. Abb. 22.<sup>22</sup> Es ergibt sich so eine Dreiteilung des Untersuchungsgebiets in einen nördlichen, einen südlichen sowie einen zentralen Cluster, zu dem auch die ehemaligen Ostgebiete gehören. In Tab. 3 sind jeweils jene Namen verzeichnet, die für die beiden Clustergrenzen (zwischen dem nördlichen und mittleren bzw. dem mittleren und südlichen Cluster) die höchsten HDI-Werte aufweisen.

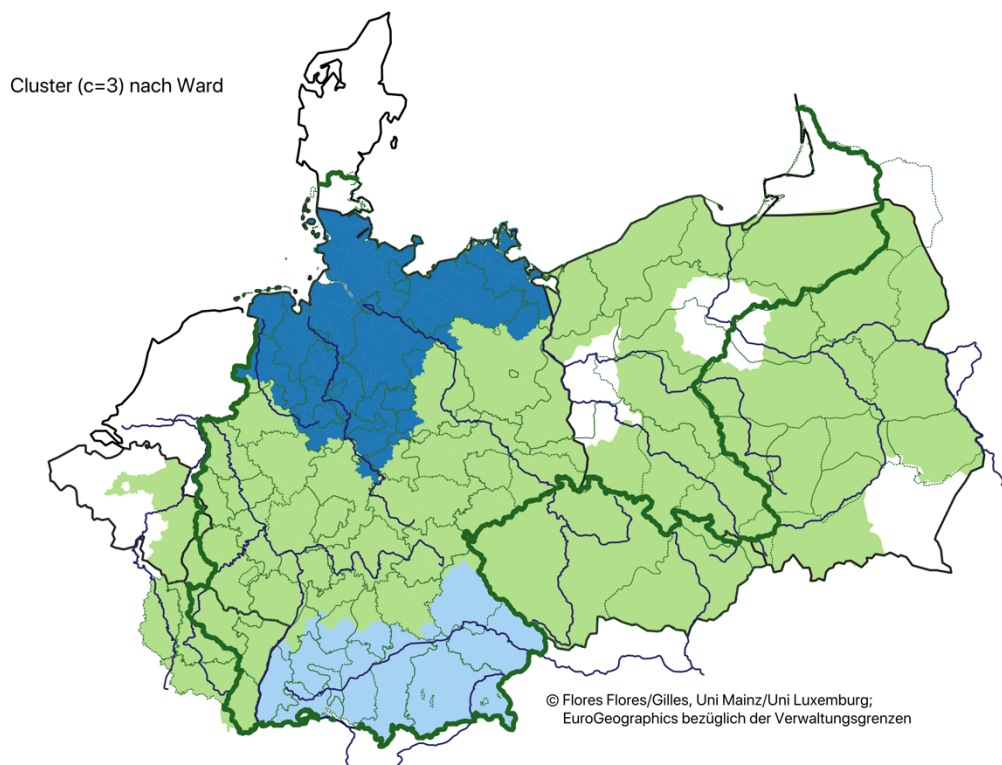


Abb. 22: Clusteranalyse nach *Ward* ( $C=3$  Cluster) (1890)

<sup>22</sup> Die gleiche Grundstruktur ergibt sich auch bei der Verwendung anderer Algorithmen zur Clusterbildung (Weighted Average, Group Average, Complete Link), und bildet sich auch in der MDS sehr klar ab.

Tab. 3: FamilienN mit maximalen HDI-Werten (*Ward*, C=3 Cluster) (1890)

Grenze Nord/Mitte				Grenze Mitte/Süd			
Cluster Nord (1)		Cluster Mitte (2)		Cluster Mitte (3)		Cluster Süd (4)	
Name	HDI	Name	HDI	Name	HDI	Name	HDI
Meyer	3617.18	Schneider	1186.47	Schmidt	1523.60	Schmid	2448.36
Schröder	1513.22	Müller	1112.71	Schulz	1144.85	Maier	2351.69
Möller	1476.78	Richter	752.12	Richter	887.25	Huber	2255.06
Petersen	1080.52	Hoffmann	735.97	Becker	868.46	Mayer	1586.83
Peters	1013.58	Wagner	631.63	Hoffmann	837.65	Bauer	1405.94
Kruse	873.52	Klein	584.73	Schulze	691.60	Meier	663.10
Hansen	837.82	Lehmann	577.29	Krause	649.47	Fischer	581.87
Voß	790.53	Schmitz	568.35	Neumann	636.29	Braun	580.51
Behrens	776.20	Hofmann	527.37	Krüger	634.65	Wagner	561.27
Ahrens	718.33	Krause	517.79	Schmitz	572.00	Lang	555.55
Meier	715.80	Schmitt	500.93	Lange	559.57	Geiger	535.22
Kröger	567.55	Weber	463.25	Müller	501.82	Beck	487.38
Martens	543.99	Scholz	462.03	Lehmann	501.65	Moser	472.90
Brandt	536.76	Wolf	425.14	Schröder	455.64	Gruber	468.60
Bartels	519.48	Bauer	402.69	Scholz	443.06	Mayr	463.01
Harms	515.77	Neumann	398.90	Schubert	402.24	Graf	461.90
Bruns	501.89	Schubert	383.56	Köhler	386.72	Wimmer	433.60
Ehlers	493.33	Schäfer	382.38	Franke	367.89	Lutz	418.39
Fricke	451.10	Weiß	374.39	Werner	328.29	Frey	409.29
Brinkmann	450.96	Braun	360.75	Schmitt	314.99	Haug	395.50
Schrader	436.86	Herrmann	347.12	Klein	311.88	Weiß	379.70
Koch	427.18	Friedrich	336.55	Otto	309.25	Haas	372.29
Schlüter	423.29	Zimmermann	330.82	Günther	287.85	Bühler	363.48
Niemann	411.41	Winkler	329.58	Meyer	284.97	Hauser	363.19
Eggers	407.39	Schmidt	321.94	Voigt	271.85	Kurz	361.93

Dass sich mit dem HDI tatsächlich die für die Clusterbildung relevanten FamilienN ermitteln lassen, zeigt sich anschaulich in Abb. 23 und Abb. 24, in denen diese jeweils 25 FamilienN mit den höchsten HDI-Werten gegenüber gestellt sind: Die FamilienN des Clusters Nord mit dem höchsten HDI an der Grenze Nord/Mitte (Spalte 1) kommen fast ausschließlich dort vor, südlich der Clustergrenze dominieren deutlich die FamilienN mit den höchsten HDI-Werten des Clusters Mitte (Spalte 2). Ebenso gilt für die FamilienN mit den höchsten HDI-Werten an der Grenze Mitte/Süd: Die Namen des Clusters Süd (Spalte 4) dominieren deutlich den südlichen Cluster, die Namen des Clusters Mitte (Spalte 3) hingegen kommen dort kaum vor, dominieren hingegen weitgehend nördlich der Clustergrenze.

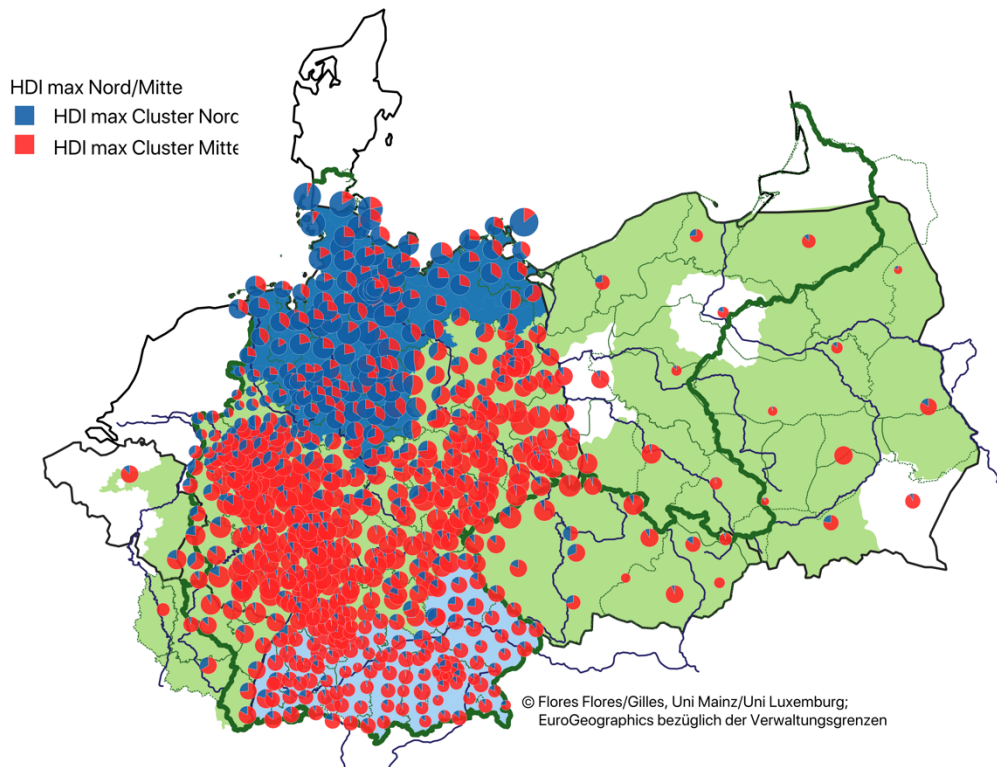


Abb. 23: relative Verbreitung der FamilienN mit dem höchsten HDI an der Grenze Nord/Mitte (Max. Symbolgröße: 6, entspricht 205,88%); Färbung der Flächen: Clusteranalyse nach *Ward* (C=3 Cluster) (1890); kartierte Namen: vgl. Tab. 3, Spalten „Grenze Nord/Mitte“, (1) und (2)

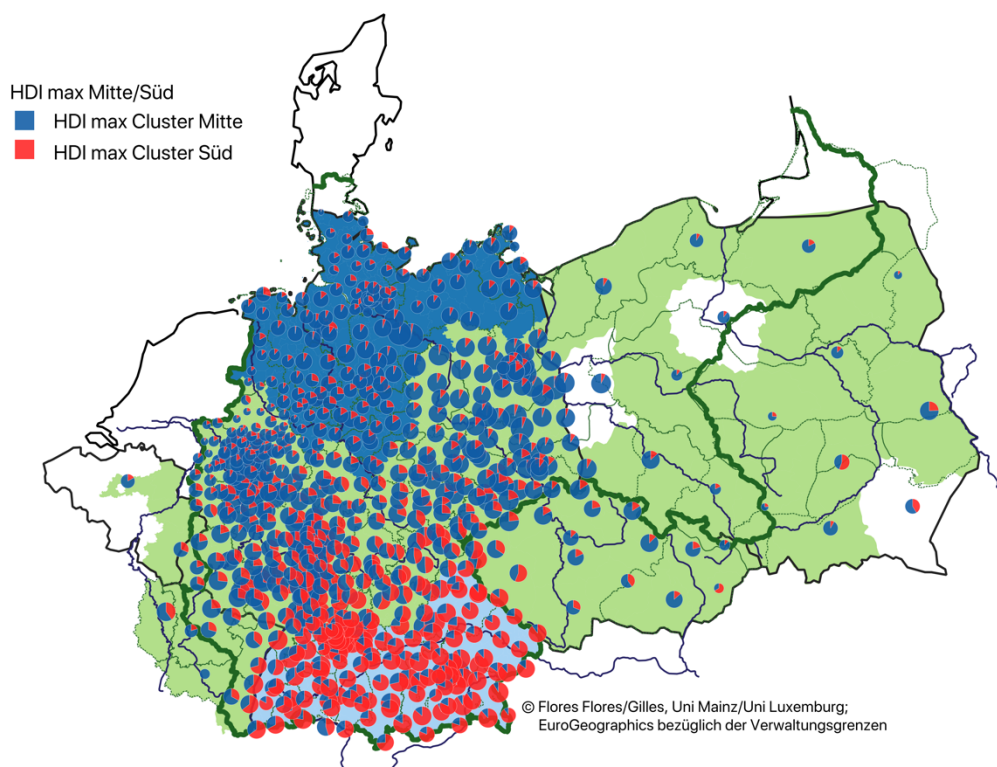


Abb. 24: relative Verbreitung der FamilienN mit dem höchsten HDI an der Grenze Mitte/Süd (Max. Symbolgröße: 6, entspricht 250,00%); Färbung der Flächen: Clusteranalyse nach *Ward* (C=3 Cluster) (1890); kartierte Namen: vgl. Tab. 3, Spalten „Grenze Mitte/Süd“ (3) und (4)

Ausgehend von dieser Erkenntnis können die Namenlisten nach solchen Namenpaaren bzw. -gruppen mit hohem HDI durchsucht werden, die sich auf beide Seiten einzelner Cluster Grenzen verteilen, um so jene konkrete, an der Raumbildung beteiligte Phänomene ermitteln zu können. Für die Grenze Mitte/Süd ist dies etwa die Schreibvarianz in den BerufsN des Typs *Meier* und *Schmidt*: Die beiden FamilienN Schmid (HDI=2448,4) und Schmidt (HDI 1523,6) finden sich auf Rang eins bzw. fünf der höchsten HDI-Werte dieser Grenze und stehen sich südlich bzw. nördlich der Cluster Grenze gegenüber, vgl. Abb. 25.

Die in Abb. 26 kartierten Namen *Maier* (HDI=2351,7/Rang 2), *Mayer* (1586,8/4), *Meier* (663,1/12) und *Mayr* (463,0/29) kommen alle fast ausschließlich südlich der Cluster Grenze vor und gehören zur TOP30 der HDI-Liste. Ihnen steht nördlich der Cluster Grenze die Schreibvariante *Meyer* gegenüber, dessen HDI mit 285,0 jedoch deutlich geringer ausfällt, aber immerhin noch für Rang 72 der TOP-Liste reicht. Der niedrigere HDI-Wert erklärt sich auch im Kartenbild: *Meyer* kommt praktisch nicht im südlichen Cluster vor, und wo doch, nicht als dominante Variante. Tatsächlich folgt das Kernverbreitungsgebiet nicht der südlichen Cluster Grenze, sondern vielmehr jener zwischen den Clustern Nord und Mitte (dort liegt *Meyer* mit einem HDI von 3617,2 auf Rang eins!) – die Variante findet sich zwar durchgehend im mittleren Cluster, allerdings nur als niederfrequente Streuung, höhere Frequenzen erreicht sie fast ausschließlich im Cluster Nord.

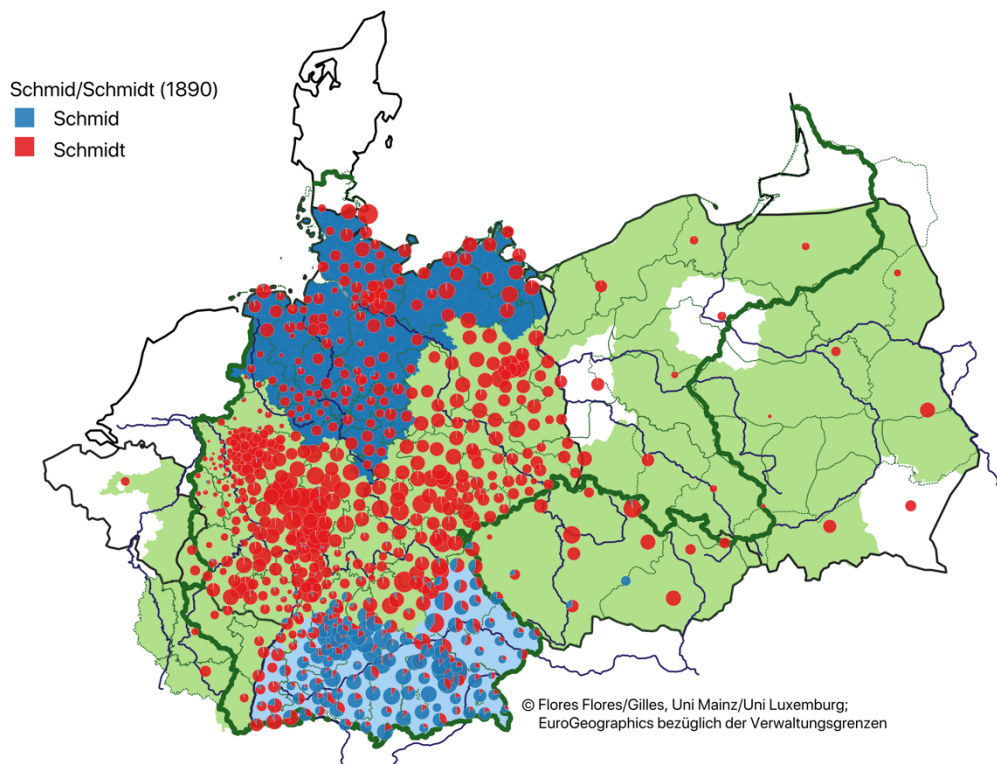


Abb. 25: relative Verbreitung der FamilienN *Schmid* vs. *Schmidt* (Max. Symbolgröße: 6, entspricht 50,69%); Färbung der Flächen: Clusteranalyse nach *Ward* (C=3 Cluster) (1890)

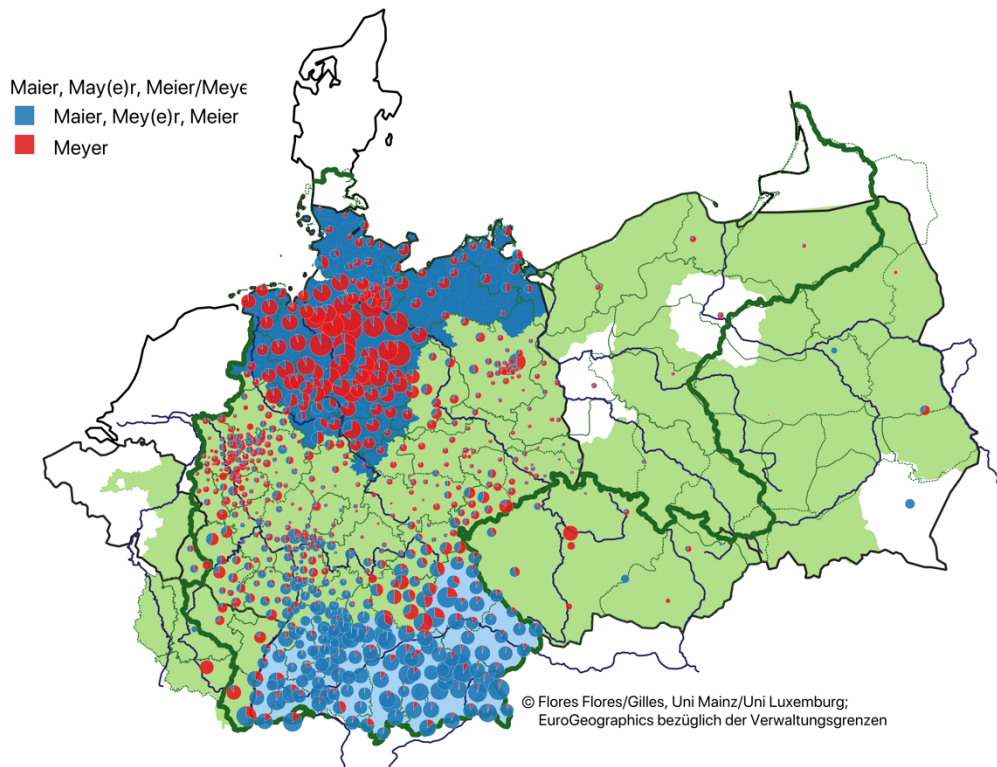


Abb. 26: relative Verbreitung der FamilienN *Maier, Mayer, Meier, Mayr* vs. *Meyer* (Max. Symbolgröße: 6, entspricht 68,01‰); Färbung der Flächen: Clusteranalyse nach *Ward* (C=3 Cluster) (1890)

## 5. Resümee

Family names are high-quality historical sources that reflect history at a high rate. To be able to use family names as historical sources, it is necessary not only to know the spread of the family names as it appears today, but also the distribution of the name before the major changes and movements of population that took place during and after World War II. (Udolph 2011: 179)

Die hier vorgestellten Verlustlisten der Soldaten des Ersten Weltkriegs stellen eine solche historische Datenquelle bislang unerreichter Qualität in Hinblick auf Größe des Inventars und Abdeckung der Gesamtbevölkerung dar. Sie sind somit ein wichtiges neues Instrument zur umfassenden Analyse der Namengeographie. Diese historischen Namendaten stellen eine valide Datenbasis dar, die die rezenten Telefonbuchdaten sinnvoll ergänzt. Sie können u.a. dazu beitragen, heterogenetisch entstandene FamilienN zu disambiguieren und vor allem den Einfluss der Migration des 20. Jh. zu überprüfen. Die hier nur cursorisch vorgestellten Analysen legen bereits nahe, dass die Verlustlisten eine weniger durch Migration gekennzeichnete Gesellschaft repräsentieren. Die über die Clusteranalyse ermittelte Dialektlandschaft der Verlustlisten erscheint daher auch deutlich konsistenter.

Die heute verfügbaren technischen Möglichkeiten erleichtern die Analyse und Kartierung großer Datensätze ungemein und die hier vorgestellte Datenbank der Verlustlisten erlaubt einen schnellen Zugang zu historischen Daten und auch das Aufdecken bislang unerkannter onymischer Muster. In Kombination mit quantitativen Ansätzen (z.B. mittels des oben vorgestellten HDI) können solche historischen Datenbanken sinnvoll für linguistisch-onomastische Fragestellungen genutzt werden.

Es ist vorgesehen, die für onomastische Zwecke aufbereitete Datenbank der Verlustlisten für die wissenschaftliche Gemeinschaft zur Verfügung zu stellen. Perspektivisch sollen auch die Verlustlisten



für Österreich/Ungarn integriert werden, die zurzeit vom Verein für Computergenealogie via Crowdsourcing digitalisiert werden. Damit werden dann grenzüberschreitende Studien etwa für die bairischen Namen ermöglicht. Es wäre darüber hinaus auch wünschenswert, dass weitere historische Datenbestände digitalisiert werden, die weite geographische Räume abdecken (z.B. Volkszählungsdaten), um punktuelle historische Tiefenbohrungen, die z.B. die sich auf die Namen eines Ortes oder einer Stadt beziehen, durch stärker (groß)raumbezogene Analysen zu ergänzen.

Schließlich eröffnet die, hier noch nicht vorgestellte, Datenbasis der (männlichen) Rufnamen ganz neue Perspektiven für die räumliche Verteilung dieser bislang vernachlässigten Namengattung.

## 6. Literatur

- Cheshire, James A., Paul A. Longley & Pablo Mateos (2009): *Family Names as indicators of Britains' Changing Regional Geography*. London (UCL Working Papers Series; 149). (<http://www.bartlett.ucl.ac.uk/casa/pdf/paper149.pdf>; 30.04.2019)
- Cheshire, James A., Paul A. Longley & Pablo Mateos (2010): Regionalisation and Clustering of Large Spatially-Referenced Population Datasets: the Case of Surnames. In R. Purves & Robert Weibel (Hrsg.), *Extended Abstracts Volume. Online Proceedings of GIScience 2010*, o.S. ([http://www.researchgate.net/profile/Pablo\\_Mateos/publication/267833644\\_Regionalisation\\_and\\_Clustering\\_of\\_Large\\_Spatially-Referenced\\_Population\\_Datasets\\_the\\_Case\\_of\\_Surnames/links/546c9c340cf2c4819f229e35.pdf](http://www.researchgate.net/profile/Pablo_Mateos/publication/267833644_Regionalisation_and_Clustering_of_Large_Spatially-Referenced_Population_Datasets_the_Case_of_Surnames/links/546c9c340cf2c4819f229e35.pdf); 30.04.2019)
- DFA = Kunze, Konrad & Damaris Nübling (Hrsg.) (2009-2018): *Deutscher Familiennamenatlas*. 7 Bde. Berlin & Boston: De Gruyter.
- Flores Flores, W. Amaru (i.Vorb.): *Die Familiennamen im historischen Luxemburgischen Raum. Eine quantitative und qualitative Analyse der onymischen Grenzen und Raumstrukturen*. (Luxemburg Studien Digital) = Flores Flores (2015): *Die Familiennamen im historischen Luxemburgischen Raum. Eine quantitative und qualitative Analyse der onymischen Grenzen und Raumstrukturen*. Diss. Uni Luxemburg.
- Flores Flores, W. Amaru & Peter Gilles (2014): Isonymie und Laskerdistanz. Möglichkeiten quantitativer Ansätze für die Familiennamengeographie in Deutschland. Vortrag: ICOS 2014, 25th International Congress of Onomastic Sciences, 25.-29.8.2014, University of Glasgow.
- Gries, Stefan Th. (2009): *Statistics for Linguistics with R. A Practical Introduction*. Berlin u.a.: De Gruyter (Trends in Linguistics. Studies and Monographs).
- Hengst, Karlheinz (2001): *Mazur und Motzki*. Slawische Familiennamen als kulturgeschichtliche Zeugen. In: Jürgen Eichhoff, Wilfried Seibicke & Miachel Wolffsohn (Hg.): *Name und Gesellschaft. Soziale und historische Aspekte der Namengebung und Namenentwicklung*. Mannheim u.a.: Dudenverlag (Thema Deutsch; 2), 209–225.
- Heuser, Rita/Dräger, Kathrin (2016): Gilles, Schillo und Jülg. Der Heiligenname Aegidius in Familiennamen in Deutschland. In: Dräger, Kathrin/Fahlbusch, Fabian/Nübling, Damaris (Hgg.): *Heiligenverehrung und Namengebung*. Berlin, Boston, S. 89-108.
- Kleiweg, Peter, John Nerbonne & Leonie Bosveld (2004): Geographic Projection of Cluster Composites. In Alan F. Blackwell, Kim Marriott & Atsushi Shimojima (Hrsg.), *Diagrammatic Representation and Inference*. Third International Conference, Diagrams 2004, Cambridge, UK, March 22-24, 2004. Proceedings. Berlin: Springer, 392–394.
- Kollmann, Cristian, Peter Gilles & Claire Muller (2016): *Luxemburger Familiennamenbuch*. Unter Mitarbeit von W. Amaru Flores Flores und Britta Weimann. Berlin/Boston: De Gruyter.

- Kunze, Konrad & Richard Kunze (2003): Computergestützte Familiennamen-Geographie. Kleiner Atlas zur Verbreitung der Apokope. *Beiträge zur Namenforschung* N.F. 38, 121–224.
- Lasker, Gabriel Ward (1985): *Surnames and genetic structure*. Cambridge u.a.: Cambridge Univ. Press (Cambridge Studies in Biological Anthropology; 1).
- Leinonen, Therese (2010): An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects. Diss., Univ. Groningen (Dissertations in Linguistics; 83).
- Leinonen, Therese (o.J.): *Gabmap Tutorial*. (<http://www.gabmap.nl/~app/doc/tutorials/tutorial/pdf/GabmapTutorial.pdf>; 30.04.2019)
- Nerbonne, John et al. (2011): GabMap – a web application for dialectology. *Dialectologia*. Special issue 2, 65–89.
- Nerbonne, John, Wilbert Heeringa & Peter Kleiweg (1999): Edit Distance and Dialect Proximity. In David Sankoff & Joseph Kruskal (Hrsg.), *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*. Stanford: CSLI, v–xv.
- Nübling, Damaris, Fabian Fahlbusch & Rita Heuser (2015): *Namen. Eine Einführung in die Onomastik*. Tübingen: Narr (Narr Studienbücher).
- Rymut, Kazimierz & Johannes Hoffmann (Hgg.) (2006): *Lexikon der Familiennamen polnischer Herkunft im Ruhrgebiet*. 1. Bd., A-L. Krakau: Verl. d. Inst. f. Poln. Sprache PAN.
- Scapoli Chiara et al. (2007): Surnames in Western Europe: a comparison of the sub-continental populations through isonymy. In *Theoretical Population Biology*, 71(1), 37–48.
- Udolph, Jürgen (2011): Familiennamen als Zeugen von Flucht, Vertreibung und Umsiedlung. In: Heuser, Rita, Damaris Nübling & Mirjam Schmuck (Hgg.): *Familiennamengeographie. Ergebnisse und Perspektiven europäischer Forschung*. Berlin & New York: de Gruyter. 179–198
- Videsott, Paul (2014): Onomatometrie als Methode der Namenforschung. In: Andrea Brendler & Silvio Brendler (Hrsg.), *Namenarten und ihre Erforschung. Ein Lehrbuch für das Studium der Onomastik*. Hamburg: Baar (Lehr- und Handbücher zur Onomastik; 1). 213–246.