

Wilfried Schiller, Aschaffenburg

## Zur Zuverlässigkeit von Häufigkeitsuntersuchungen für Familiennamen aus Telefonverzeichnissen

In jüngster Zeit wurde die Häufigkeit von Familiennamen (FN) in geographischer Aufteilung verstärkt untersucht, vermutlich nicht zuletzt deshalb, weil mit Personalcomputern (PC) der erforderliche Rechenaufwand wesentlich verringert werden konnte. Mit Telefonverzeichnissen auf elektronischen Speichermedien (CD, Internet) entfällt zudem die aufwendige Datenerhebung. Nur selten wird jedoch danach gefragt, wie zuverlässig denn die Ergebnisse solcher Untersuchungen sein können. Das soll hier geschehen, denn die Verteilungskarten von FN sind nicht nur aus sprachwissenschaftlicher Sicht, sondern auch für den Genealogen interessant.<sup>1</sup>

So stellen Konrad und Richard KUNZE ein „nicht öffentlich zugängliches Programm“ für computergestützte Verteilungskarten von FN vor,<sup>2</sup> das auf nicht näher beschriebenen Telefonverzeichnissen des Jahres 1996 basiert. Mittels dieser Software entstanden alle Berechnungen und Karten des „dtv-Atlas Namenkunde“.<sup>3</sup> Im „DUDEN Familiennamen“ geben Rosa und Volker KOHLHEIM die CD-ROM „D-Info '97“ (Mannheim 1997, Stand der Daten 1996) an,<sup>4</sup> ohne dass erläutert wird, wie daraus z. B. die Tabelle der 200 häufigsten FN und die Deutschlandkarte der regional häufigsten FN entstanden sind. Jürgen UDOLPH arbeitet mit der Telefon-CD „DT-Info & Route“ von 1998<sup>5</sup> sowie mit dem *geogen* genannten Internet-Angebot von Christoph STÖPEL.<sup>6</sup>

Letzteres zeigt in den Grenzen der deutschen Bundesländer für jeden Stadt- oder Landkreis an, zu welcher von sechs vorgegebenen Häufigkeitsklassen der gesuchte FN in diesem Gebiet ge-

hört. Positiv an dieser Darstellung ist, dass die Namenhäufigkeit sowohl direkt als auch auf die Bevölkerungszahl des Kreises bezogen werden kann, also neben absoluten auch relative Häufigkeiten abgebildet werden. Außerdem wird angegeben, auf welchem Rangplatz der Häufigkeit der Name steht und in welchen Kreisen Maximal- bzw. Minimalwerte zu finden sind. Weniger günstig ist die feste Einteilung in stets sechs Häufigkeitsklassen, die eine differenzierte Aussage nur für Namensvorkommen in „Nahe-Null-Bereich“ erlaubt.

Unsere Arbeiten<sup>7,8,9</sup> förderten in verschiedenen *klicktel*-Versionen eine Reihe von Erfassungsfehlern zutage, die sich von Ausgabe zu Ausgabe fortpflanz(t)en. Dabei wurden nur drei Datenelemente ausgewertet: Postleitzahl (PLZ), Ort und Familienname. Zur Fehlerbehandlung, nicht zur Datenanalyse, wurden zusätzlich die Vornamen selektiert sowie das Datenelement „Anrede“ zur Erkennung von Firmeneinträgen.

Ortsbezeichnungen – bei ihnen kommen die wenigsten Fehler vor – wurden ungeprüft übernommen. Ignoriert wurden dagegen Datensätze mit PLZ, die zum Ausgabezeitraum der *klicktel*-CD von der Deutschen Post nicht belegt waren. Gegenüber den erfassten FN ist allerdings in dreifacher Hinsicht eine kritische Haltung angebracht:

1. Es gibt an die hundert *Baumann, Beckmann, Bergmann, Brinkmann, Edelmann, Engelmann, Großmann, Hermann, Herrmann, Hoffmann, Hofmann, Kellermann, Lehmann, Naumann, Neumann, Reichmann, Wichmann, Zimmermann* usw., die am Ende mit *-nnn* (dreifachem *-n!*) erscheinen.
2. Auch *Schramm* mit *-mmm* oder *Schröter* mit *-rr-* sind wohl falsch erfasst worden.
3. FN, die Sonderzeichen enthalten, wie Punkt, Komma, Kaufmanns-Und (&), Schrägstriche – wohlgemerkt alles bei FN, nicht etwa bei Firmeneinträgen – dürften wohl in einem deutschen Standesamt zu Problemen führen. Gerade diese Gruppe bildet mit Abstand die größte Fehlermenge.<sup>10</sup>

Die hier aufgeführten Fehler in den FN lassen sich relativ leicht erkennen. Zweifellos können ungewollte Verdoppelungen von Konsonanten, hinzugefügte oder weggelassene Buchstaben aber auch bei selteneren Namen entstanden sein, und wer weiß dann, ob sie so oder anders geschrieben sein müssten. Quantifizieren lässt sich die Fehlerrate bestenfalls an Einzelbeispielen, wenn beispielsweise wirkliche Bevölkerungsdaten zur Verfügung stehen, die auch die Grundlage von Wählerverzeichnissen sind und daher besser gepflegt werden als Telefonverzeichnisse. Nach unseren Untersuchungen<sup>8, 9</sup> liegt die Fehlerquote aus der Datenerfassung bei FN, PLZ und Ort auf der *klicktel*-CD insgesamt bei zwei Prozent.

Jedes Telefonverzeichnis auf CD hat seine Stärken und seine Schwächen. Stärken sind meist leicht zu erkennen, bei den Schwächen ist das bisweilen schwieriger. So verfügt *klicktel* über einen Filter, der zunächst keinerlei Verdacht erregt. Das Problem soll am typisch bayerischen Namen *Huber* erläutert werden. Sucht man diesen im PLZ-2-Gebiet<sup>11</sup> 80 (München), dann werden von *klicktel* nur fünf Treffer angezeigt, wenn nach *vollständigem* Namen gesucht wird. Lässt man dagegen alle mit *Huber* beginnenden Namen suchen, dann gibt es plötzlich 1 359 Treffer, davon 1 288 reine *Huber*, 3 *Huber* mit Komma, 2 *Huber* mit Bindestrich sowie insgesamt 56 *Hubert*, *Huberth*, *Huberti*, *Hubertus* und *Hubertz*. Weshalb bei vollständiger FN-Suche die Trefferquote von 1 288 auf fünf *Huber*, also auf 0,4 Prozent der erfassten Daten zu diesem FN sinkt, das bleibt ein Geheimnis des *klicktel*-Programms. Jeder Nutzer sollte von dieser Schwäche wissen und immer den Filter „Name unvollständig“ setzen. Bei dieser Filterstellung hat man dann die Möglichkeit, auch die FN mit irrtümlich erfassten Punkten, Schrägstrichen, dreifachen Endbuchstaben, z. B. bei den *-mann* usw., zu akzeptieren, um sie anschließend zu korrigieren. Gleiches gilt für die Doppelnamen (mit Bindestrich).

Ein weiterer Filter ist die Art des verzeichneten Kommunikationsgerätes. *Klicktel* bietet hierfür an: *Alle*, *Telefon*, *Telefax*, *Mobil* oder *Internet*. Wer seine Faxnummer oder die Internetadresse ins

Telefonverzeichnis aufnehmen lässt, der ist mit Sicherheit auch mit der Telefonnummer verzeichnet und würde – ungefiltert – doppelt gezählt. Die Anzahl eingetragener Handynummern ist vergleichsweise gering. Hier könnte es zu Doppelzählungen kommen, wenn zum gleichen Namen ein Festnetzanschluss eingetragen ist. Andererseits gibt es ganz mobile, sicher vor allem jüngere, Leute, die gar kein Festnetztelefon mehr betreiben. Die Häufigkeitsverhältnisse beider Gruppen schwanken auch regional besonders stark. 90 bis 95 Prozent der Verzeichniseinträge werden durch (Festnetz)Telefon repräsentiert. Bei Berücksichtigung weiterer Kommunikationsarten besteht die Gefahr von Doppelzählungen.

Neben der Fehlerbehandlung, die ja auch schlicht im Ignorieren bestehen kann, und den Auswertungsfiltren gibt es beim Datenexport aus Telefonverzeichnissen mindestens fünf Probleme. Je nachdem, wie man sich entscheidet, erhält man unterschiedliche Ergebnisse:

Problem *Firmen-Einträge*: Exportiert man Daten aus *klicktel*, dann erhält man auch das Datenelement „Anrede“ mit vier möglichen Belegungen: *Firma, Frau, Herr* oder *kein Inhalt*. Insbesondere Gewerbe- und Handelsbetriebe firmieren oft unter dem FN der Gründer oder Inhabers der Firma. Der Gründer kann längst nicht mehr unter den Lebenden sein, der Inhaber hat meist auch einen privaten Telefonanschluss. Deshalb erscheint es zweckmäßig, alle als solche gekennzeichneten Firmeneinträge von der Weiterverarbeitung auszuschließen.<sup>12</sup> Vor allem unter den Datensätzen mit fehlendem Inhalt bei „Anrede“ sind oftmals Vereine, Organisationen, Verbände, Institutionen, Kirchen usw. aufgeführt. Nach ihrem Selbstverständnis wollen deren Vertreter durchaus nicht als „Firma“ bezeichnet werden, gehören aber auch nicht zu den FN.

Problem *unvollständige PLZ*: Neben den absolut telefonbuchscheuen Mitmenschen, die zwar kommunizieren, aber nicht von jedermann angerufen werden wollen, gibt es relativ viele, die zwar den FN, nicht aber die komplette Anschrift angeben. Wird dabei auch die PLZ nicht vollständig genannt (Musterbeispiel Hamburg

mit großem Anteil PLZ 2...), entstehen für fein gegliederte Auswertungen Probleme. Immerhin finden sich unter *Müller* 5921 Namenträger, die keine vollständige PLZ vermerken ließen. Wir verfahren bisher so, dass für jede Stufe der Auswertung die identifizierbaren Teile der PLZ akzeptiert wurden. Dadurch erhält man bei PLZ-1-Auswertungen höhere Werte als bei PLZ-2 usw.

Problem *FN mit Umlauten und mit ss bzw. ß* in zentraler bzw. finaler Position: Die Telefonverzeichnisse auf CD ermöglichen eine schnelle Suche nach einzelnen Eintragungen. Dabei ist es vorteilhaft, die bei FN immer etwas ungewisse Schreibweise der Umlaute (*ä, ae; ö, oe; ü, ue*) und des *ss* bzw. *ß* im gleichen Suchlauf zu finden. Ist man aber nicht an einem einzelnen Namenträger interessiert, sondern an allen, dann muss man die *Müller*, die *Vofß* usw. in ihren unterschiedlichen Schreibweisen jeweils zusammengefasst hinnehmen. KUNZE<sup>12</sup> hat das entsprechend gekennzeichnet, „DUDEN Familiennamen“ nicht, aber dennoch so gearbeitet. STÖPEL<sup>13</sup> dagegen unterscheidet richtigerweise und zählt bundesweit 262 720 *Müller* mit *-ü-* und 1 615 mit *-ue-*, 62 026 *Schäfer* mit *-ä-* und 4 552 mit *-ae-*, 11 499 *Vofß* mit *-ß* und 5 877 *Voss* mit *-ss* usw. Die Schlussfolgerung kann nur lauten, sich dem stöpelschen Verfahren anzuschließen, denn bei den diversen *Schmied/Schmidt/Schmitt ...* und *Maier/Mayer/Meier/Meyer ...* oder *Franck/Frank* wird ja auch korrekt nach jeder einzelnen Schreibweise unterschieden.

Problem *Doppelnamen*: Der Anteil der verzeichneten Doppelnamen ist wiederum regional sehr unterschiedlich. In „DUDEN Familiennamen“ ist nicht angegeben, ob in den Häufigkeitsübersichten auch die Doppelnamen berücksichtigt sind. KUNZE dagegen schreibt, dass „versucht wurde, Doppelinträge (Privat- und Geschäftsanschlüsse), Doppelnamen (*Müller-Erb*) usw. möglichst zu eliminieren.“ Aus dem Kontext der Doppelinträge kann geschlossen werden, Doppelnamen würden bei KUNZE weggelassen. Das scheint uns – bei der Vielzahl ihres Auftretens – nicht gerechtfertigt zu sein, zuviel Datenmaterial bliebe dann ungenutzt. Eine Auflösung in beide Namenbestandteile bläht jedoch die Anzahl unzu-

lässigerweise auf. Wir haben deshalb bei Doppelnamen jeweils nur den *ersten* Teil berücksichtigt, den zweiten aber ignoriert.

Problem *Namenszusätze*: Hier gilt faktisch dasselbe wie für Doppelnamen. Namenszusätze wie die seit 1918 abgeschafften Adelstitel, obwohl sie rechtlich ja keine mehr sind, sondern nur Namenbestandteile, werden ignoriert.<sup>14</sup> Aus den *von Müller* werden also schlichte *Müller* usw. Ganz andere Namenszusätze müssen aber auch noch erkannt und behandelt werden. Das betrifft die in Telefonverzeichnissen gar nicht seltenen Zusätze *junior* und *senior* in den unterschiedlichsten Schreibweisen oder Abkürzungen. Auch Berufsbezeichnungen wie *Architekt*, *Steuerberater*, *Notar* usw. finden sich als Namenszusätze, von den vielen akademischen Titeln unterschiedlichster Grade ganz zu schweigen. Filtert man solche Namenszusätze nicht heraus, dann wird mit jedem ein zusätzlicher FN erkannt – mit entsprechenden Folgen.

Es mag verwundern, dass bei den aufgezählten Erfassungsfehlern und deren Behandlung, der angewendeten Filter und der Art und Weise, wie den o.g. fünf Problemen begegnet wird, die Bezugsjahre der Telefonverzeichnis-CD nicht genannt worden sind. Diesbezüglich enthält die Meinung „Je mehr Handys mit der Zeit, desto weniger Festnetzanschlüsse; je weniger Festnetzanschlüsse, desto weniger Verzeichniseintragungen“ (J. UDOLPH) zweifellos viel Wahres. Angesichts dessen könnte nun angenommen werden, dass die Aussagen der verschiedenen Autoren nicht mehr vergleichbar sind. Dem ist nicht so, wie die nachfolgende Tabelle zeigt.<sup>15</sup>

Die hoch signifikanten Korrelationen zwischen den Werten in den vier Datenspalten der 50 häufigsten deutschen FN belegen nahezu identische Aussagen: Die niedrigste Korrelation mit „nur“ 0,9982 besteht zwischen den Resultaten des „DUDEN Familiennamen“ und unserer Darstellung, die höchste mit 0,9993 zwischen dem „DUDEN Familiennamen“ und STÖPEL.

Auch bei der Darstellung der häufigsten deutschen FN je PLZ-2-Gebiet nach dem Stand von 2003 zeigt sich, dass mit ganz gerin-

Name	KUNZE	KOHLHEIM	SCHILLER	STOEPER
Müller -ue-	1 (270,0)	1 (324,1)	1 (274,0)	1 (264,3)
Schmidt	2 (194,9)	2 (235,8)	2 (201,7)	2 (194,0)
Schneider	3 (114,8)	3 (142,1)	3 (117,8)	3 (116,8)
Fischer	4 (99,2)	4 (122,9)	4 (97,8)	4 (99,0)
Meyer	5 (92,9)	5 (106,4)	5 (96,4)	6 (84,4)
Weber	6 (85,5)	6 (106,2)	6 (84,8)	5 (86,6)
Schulz	7 (78,2)	9 (93,3)	10 (70,6)	8 (75,5)
Wagner	8 (77,5)	7 (97,2)	7 (79,7)	7 (80,1)
Becker	9 (76,7)	8 (93,4)	8 (74,0)	9 (74,9)
Hoffmann	10 (72,8)	10 (89,4)	9 (73,9)	10 (72,2)
Koch	12 (60,7)	12 (75,1)	13 (62,6)	13 (60,8)
Bauer	13 (59,5)	13 (74,0)	12 (62,9)	14 (59,2)
Schröder -oe-	14 (57,3)	16 (63,9)	16 (56,4)	15 (55,1)
Klein	15 (56,9)	15 (67,9)	15 (56,6)	16 (54,3)
Richter	16 (56,1)	14 (72,5)	14 (60,6)	12 (61,0)
Wolf	17 (50,3)	17 (63,3)	17 (53,2)	17 (52,0)
Neumann	18 (48,6)	18 (59,4)	18 (49,2)	18 (48,4)
Schwarz	19 (45,4)	19 (55,1)	19 (45,8)	19 (44,5)
Schmitz	20 (44,0)	24 (50,8)	26 (41,6)	28 (39,7)
Krüger -ue-	21 (43,9)	21 (53,4)	24 (42,2)	21 (43,4)
Braun	22 (43,2)	22 (52,7)	21 (44,4)	22 (42,5)
Zimmermann	23 (42,6)	20 (53,5)	20 (44,6)	20 (43,7)
Schmitt	24 (41,7)	27 (50,0)	23 (43,1)	26 (40,1)
Lange	25 (41,2)	26 (50,2)	27 (41,0)	25 (40,7)
Hartmann	26 (41,2)	25 (50,7)	25 (42,1)	24 (40,7)
Hofmann	27 (40,7)	23 (51,1)	22 (43,9)	23 (42,1)
Krause	28 (40,1)	29 (48,8)	29 (40,1)	29 (39,6)
Werner	29 (39,2)	28 (48,8)	28 (40,4)	27 (39,9)
Meier	30 (38,9)	30 (46,6)	30 (39,8)	32 (37,6)
Schmid	31 (37,0)	31 (45,8)	34 (36,5)	34 (35,7)
Schulze	32 (36,7)	33 (45,1)	41 (33,7)	33 (35,8)
Lehmann	33 (35,4)	32 (45,2)	32 (37,7)	30 (37,9)
Köhler -oe-	34 (34,4)	35 (42,6)	35 (36,0)	35 (35,4)

Name	KUNZE	KOHLHEIM	SCHILLER	STOEPEL
<i>Maier</i>	35 (34,2)	34 (42,7)	33 (37,4)	38 (34,3)
<i>Herrmann</i>	36 (34,1)	36 (42,2)	37 (35,0)	37 (34,3)
<i>König -oe-</i>	37 (33,9)	38 (41,4)	36 (35,4)	36 (34,5)
<i>Mayer</i>	38 (33,9)	39 (41,0)	39 (34,6)	40 (32,3)
<i>Walter</i>	39 (33,5)	37 (41,5)	38 (34,7)	39 (33,7)
<i>Peters</i>	40 (32,2)	43 (38,5)	44 (31,7)	45 (30,6)
<i>Möller -oe-</i>	41 (32,2)	44 (37,5)	46 (31,3)	44 (30,7)
<i>Huber</i>	42 (31,9)	40 (39,4)	40 (34,6)	42 (31,2)
<i>Kaiser</i>	43 (31,6)	41 (39,3)	42 (33,1)	41 (31,7)
<i>Fuchs</i>	44 (31,5)	42 (39,0)	43 (32,8)	43 (31,2)
<i>Scholz</i>	45 (30,7)	45 (37,4)	47 (30,9)	46 (30,1)
<i>Weiss -ß</i>	46 (29,8)	47 (36,7)	31 (38,1)	31 (37,6)
<i>Lang</i>	47 (29,8)	46 (36,8)	45 (31,7)	47 (29,8)
<i>Jung</i>	48 (28,6)	48 (35,1)	48 (29,7)	48 (28,3)
<i>Hahn</i>	49 (26,1)	49 (33,4)	49 (28,2)	49 (27,5)
<i>Keller</i>	50 (26,1)	53 (32,4)	50 (27,7)	53 (26,3)

Tab. 1: Häufigste deutsche FN laut verschiedenen Autoren

gen Abweichungen dasselbe Resultat erzielt wurde wie bei der Häufigkeitskarte im „DUDEN Familiennamen“ mit Stand von 1996 (ähnlich auch bei KUNZE, aber nur für den nördlichen Teil Deutschlands). In nur sieben der 95 PLZ-2-Gebiete gibt es zum Teil ganz geringe Unterschiede. Meist differiert das Ergebnis nur um eine Person oder um wenige Namenträger. Verschiedene Bezugsjahre für die Verzeichnis-Daten bewirken wesentlich geringere Unterschiede als solche aus den verschiedenen methodischen Ansätzen.

Bei so viel Übereinstimmung erhebt sich die Frage, weshalb überhaupt die Zuverlässigkeit von Häufigkeitsuntersuchungen von FN aus Telefonverzeichnissen diskutiert werden soll.

Gleichgültig, ob eigene Programme entwickelt wurden oder auf Standardsoftware zurückgegriffen werden muss – die Resul-



tate können aus verständlichen Gründen nicht so exakt sein wie etwa die aus Volkszählungen gewonnenen, denn

- nicht jeder hat ein Telefon, schon gar nicht in vielköpfigen Familien,
- nicht jeder, der ein Telefon hat, lässt seine Telefonnummer veröffentlichen,
- und nicht jeder, der im Telefonverzeichnis vorkommt, ist dort mit auswertbarer PLZ registriert,
- mancher ist dagegen im Telefonverzeichnis als Privatperson und als Firmeninhaber oder als Nutzer von Festnetztelefon, Telefax oder Handy mehrfach verzeichnet.

So erhebt sich die Frage, ob Telefonverzeichnisse grundsätzlich als statistisch zuverlässige Stichprobe für die Gesamtbevölkerung zu gebrauchen sind.

Für KUNZE ist klar: „Telefonverzeichnisse bieten eine namenkundliche Datenbasis ersten Ranges. Sie wird hier erstmals umfassend für Deutschland genutzt ...“<sup>16</sup> Am zeitlichen Primat gibt es keinen Zweifel, denn schon früher hatte sich KUNZE mit Telefonverzeichnissen als Grundlage für Namensverhältnisse in der Gesamtbevölkerung befasst. Anfangs ging er davon aus, dass im Mittel jedem Telefonverzeichniseintrag *drei* reale Menschen entsprechen. Nach den Fortschritten bei Telefonanschlüssen im östlichen Deutschland sind es nach seiner Meinung nun „durchschnittlich ca. 2,8 Träger des betr. Namens ...“.

Hier soll zunächst die hypothetische Existenz eines solchen Faktors<sup>17</sup> überprüft werden. Multipliziert man die von KUNZE genannte Gesamtzahl von 28,448 Millionen Telefonverzeichniseintragungen mit dem Faktor 2,8, so erhalten wir für 1995 eine Gesamtbevölkerung Deutschlands von 79,655 Millionen über alle 960 000 verschieden geschriebenen FN. Die Berichte des Statistischen Bundesamtes Wiesbaden nennen für 1995 den Wert 81,817 Millionen – demnach wäre der Faktor insgesamt 2,876, also eher 2,9 als 2,8. Ob die Abweichung zu groß ist (immerhin steckt dahinter eine absolute Größe von mehr als zwei Millionen Bewohner Deutschlands

mehr oder weniger!), das soll hier nicht weiter geprüft werden. Vielmehr ist zu fragen, ob sich die Abweichungen für die einzelnen FN in statistisch vertretbarer Weise um einen Mittelwert von 2,8 bis 2,9 bewegen. Exakt ließe sich diese Frage nur durch Auswertung von Volkszählungsdaten, Wähler- oder Einwohnermeldeamtsverzeichnissen beantworten.

Für zwei Orte mit 2500 bzw. 69000 Einwohnern<sup>18</sup> standen exakte Bevölkerungszahlen je FN zur Verfügung. Sie konnten unmittelbar mit den *klicktel*-Werten verglichen werden. Beide liefern keine Bestätigung für die Existenz des von KUNZE angenommenen Faktors (im Mittel 2,41 bzw. 2,68). Noch gravierender wird die Variationsbreite bei Betrachtung einzelner FN.<sup>19</sup> Das zeigt die Abbildung für die häufigsten FN in Aschaffenburg.

Der Vergleich der realen Bevölkerungsdaten mit den *klicktel*-Eintragungen lässt folgende Schlüsse zu:<sup>20</sup>

1. Grundsätzlich können die in Telefonverzeichnissen eingetragenen FN als repräsentativ für die Bevölkerungszahlen gelten – die geordneten Werte aus *klicktel* haben einen ähnlichen Verlauf wie die der Einwohner, aber er ist nur *ähnlich*, keinesfalls identisch.
2. Es gibt allerdings keinen einheitlichen Umrechnungsfaktor. Die Werte 2,8 bzw. 3,0 treten in den behandelten Stichproben bei den untersuchten häufigen FN eher selten auf.
3. Die realen Werte der Einwohnerzahlen je FN schwanken für die Anzahl der Telefonbucheinträge im Bereich zwischen 1,9 und 3,0 mit ganz deutlichen Abweichungen nach oben bzw. unten, vor allem bei selteneren Namen.
4. Die Telefonverzeichnisdaten bilden keine zuverlässige Stichprobe für die Gesamtbevölkerung bei der Häufigkeitsermittlung der darin erfassten FN.

Mit diesen Schlussfolgerungen kann nur zu großer Vorsicht geraten werden, wenn es darum geht, von der Anzahl der Telefonverzeichniseintragungen auf die realen Bevölkerungswerte zu schließen.

KUNZE meint,<sup>21</sup> die Telefonverzeichnisse wären eine „namenkundliche Datenbasis ersten Ranges.“ Sie „eignen sich ausgezeich-

net für namenssystematische und -geographische Untersuchungen. Es steht auch keine bessere Alternative zur Verfügung.“ Wir können – leider – nur dem letzten zitierten Satz vollinhaltlich zustimmen.

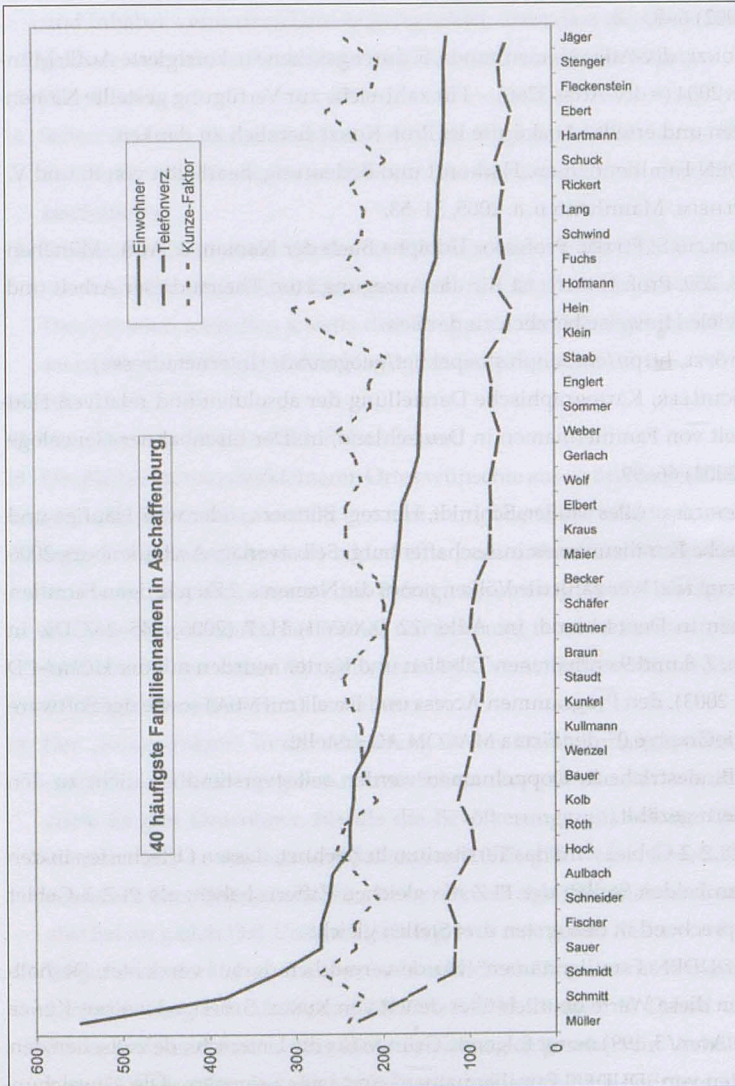


Abb. 1: 40 häufigste FN in Aschaffenburg

## Anmerkungen

- 1 VERFASSER, promoviert, Familienforscher mit besonderem Interesse für angewandte Statistik und Kartographie, ist Eisenbahner im Ruhestand.
- 2 K. KUNZE, R. KUNZE, Hohe Richter-Dichte in Sachsen, in: Computergenealogie 1 (2002) 6–8.
- 3 K. KUNZE, dtv-Atlas Namenkunde, 5. durchgesehene u. korrigierte Aufl., München 2004 (= dtv-Atlas 3266). – Für zahlreiche zur Verfügung gestellte Namenskarten und erteilte Auskünfte ist Prof. KUNZE herzlich zu danken.
- 4 DUDEN Familiennamen. Herkunft und Bedeutung, bearbeitet von R. und V. KOHLHEIM, Mannheim u. a. 2005, 51–53.
- 5 J. UDOLPH, S. FITZEK, Professor Udolphs Buch der Namen, 3. Aufl., München 2006, 259. Prof. UDOLPH ist für die Anregung zum Thema dieser Arbeit und für viele Hinweise herzlich zu danken.
- 6 C. STÖPEL, <http://christoph.stoepel.net/geogen/v3/> (Internetadresse).
- 7 W. SCHILLER, Kartographische Darstellung der absoluten und relativen Häufigkeit von Familiennamen in Deutschland, in: Der Eisenbahner-Genealoge 31 (2004) 66–99.
- 8 W. SCHILLER, Alles Müller, Schmidt, Herzog, Büttner ... oder wie? Häufige und typische Familiennamen in Aschaffenburg, Selbstverlag, Aschaffenburg 2006.
- 9 W. SCHILLER, Wer zählt die Völker, nennt die Namen ...? Zu häufigen Familiennamen in Deutschland, in: Adler 22 (XXXVII) H. 7 (2006) 245–267. Die in Anm. 7, 8 und 9 enthaltenen Tabellen und Karten wurden mit der *klicktel*-CD (Juli 2003), den Programmen Access und Excel (mit VBA) sowie der Software „RegioGraph 6.0“ der Firma MACON AG erstellt.
- 10 Die Bindestriche in Doppelnamen werden selbstverständlich nicht zu den Fehlern gezählt.
- 11 Als PLZ-2-Gebiet wird das Territorium bezeichnet, dessen Ortschaften in den ersten beiden Stellen der PLZ die gleichen Ziffern haben; als PLZ-3-Gebiet entsprechend in den ersten drei Stellen gleich.
- 12 Im „DUDEN Familiennamen“ wurde vermutlich darauf verzichtet. Deshalb liegen diese Werte deutlich über denen von KUNZE, STÖPEL oder VERF. KUNZE (wie Anm. 3, 199) nennt folgende Gründe für die Unterschiede zwischen den Werten von „DUDEN Familiennamen“ und seinen eigenen: „Die Abweichun-

- gen hängen ... schließlich damit [zusammen], dass ... versucht wurde, Doppelseinträge (Privat- und Geschäftsanschlüsse), Doppelnamen (Müller-Erb) usw. möglichst zu eliminieren." VERF. ist an solchen Versuchen bei den Doppelseintragungen gescheitert, so dass es interessant wäre, woran sie erkannt werden sollen. Die einzige Möglichkeit wäre wohl die gleiche Anschrift von Firma und Inhaber – eine unsichere Angelegenheit, denn dass die Chefs im Firmengelände wohnen, ist heutzutage wohl eher nicht mehr die Regel.
- 13 Wie Anm. 6.
  - 14 Selten sind die Namenszusätze *von* nicht, insgesamt finden sich 3445 Graf/Gräfin/Freifrau/Freiin/Baron/Baronin/Baroness *von* [...] in den *klicktel*-Verzeichnissen.
  - 15 Tabelle mit gerundeten Werten in 1000 aus Anm. 9, 251–252, teilweise übernommen. Die Überschrift KOHLHEIM steht für „DUDEN Familiennamen“. Die Datenspalten enthalten jeweils den Rangplatz der Häufigkeit und in Klammern die auf 1000 gerundete Häufigkeit des jeweiligen FN.
  - 16 Wie Anm. 3, 199.
  - 17 Vom VERF. als „KUNZE-Faktor“ bezeichnet.
  - 18 Der Bürgermeister des kleineren Ortes wünschte aus Gründen des Datenschutzes, dass der Ortsname nicht genannt wird. Die 69000 ständig in Aschaffenburg Wohnenden (Anm. 8) bilden eine ganz ordentliche Stichprobe, wenn man zum Vergleich etwa an den Befragungsumfang für die Wahlprognosen in Deutschland denkt, der immer nur bei rund tausend Personen liegt. Die Aschaffenburger haben übrigens real 14600 verschiedene FN, im Mittel also 4,7 Träger desselben Familiennamens.
  - 19 Der „KUNZE-Faktor“ in der Abbildung ist aus Darstellungsgründen mit 100 multipliziert wiedergegeben. Seine Schwankungsbreite hat folgende Ursachen: Es gibt Einwohner, für die die Bevölkerungszahl mit der *klicktel*-CD übereinstimmt – das wäre ein Faktor von 1,0. Es gibt weiterhin z. B. 13 Träger desselben Familiennamens, für die nur ein einziger Verzeichniseintrag steht, also Faktor gleich 13,0. Und schließlich gibt es eine ganze Reihe von FN, deren Träger zwar dort leben, die aber auf der *klicktel*-CD gar nicht vorkommen, in Aschaffenburg z. B. 15 *Kathrein*, 14 *Schmick*, 13 *Deobald*, 12 *Debald* und je 11 *Berber*, *Janke* und *Stenzel*. Das wäre mathematisch gesehen ein Verhältnis von  $x$  zu Null, also das, was die Mathematiker als unendlich definieren.

- 20 Um die Ergebnisse von Aschaffenburg verifizieren zu können, wurden auch noch die realen Daten der Stadt Hof in Bayern untersucht (Die Arbeit liegt inzwischen gedruckt vor: W. SCHILLER, Alles Müller, Schmidt, Schiller, Fichtner ... oder wie? – Häufige und typische Familiennamen in Hof, in: *Miscellanea curiensis*. Beiträge zur Geschichte und Kultur Nordoberfrankens und angrenzender Regionen; 56. Bericht (2008) 79–98). Auch in Hof kann ein einheitlicher Umrechnungsfaktor nicht nachgewiesen werden.
- 21 Wie Anm. 16.