

SOEPpapers

on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel study at DIW Berlin

990-2018

Die Scrabble-Score-Methode zur Messung sprachlicher Komplexität – Ein Test anhand von 90.000 Rufna- men aus dem SOEP

Anna-Maria Balbach

SOEPPapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPPapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPPapers are available at
<http://www.diw.de/soeppapers>

Editors:

Jan **Goebel** (Spatial Economics)
Stefan **Liebig** (Sociology)
David **Richter** (Psychology)
Carsten **Schröder** (Public Economics)
Jürgen **Schupp** (Sociology)

Conchita **D'Ambrosio** (Public Economics, DIW Research Fellow)
Denis **Gerstorff** (Psychology, DIW Research Fellow)
Elke **Holst** (Gender Studies, DIW Research Director)
Martin **Kroh** (Political Science, Survey Methodology)
Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Fellow)
Thomas **Siedler** (Empirical Economics, DIW Research Fellow)
C. Katharina **Spieß** (Education and Family Economics)
Gert G. **Wagner** (Social Sciences)

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: soeppapers@diw.de



Die Scrabble-Score-Methode zur Messung sprachlicher Komplexität –
Ein Test anhand von 90.000 Rufnamen aus dem SOEP

von Anna-Maria Balbach

Münster, im Oktober 2018

Abstract

For several years, American studies of different disciplines have established the so-called Scrabble score method as a measure of the linguistic complexity of names. This method scores each name according to defined rules, similar to the popular board game Scrabble, in which each word receives a score defined by the sum of the numerical values of their characters. The numerical Scrabble scores allow comparing and analyzing names in flexible and quantitative ways.

In this study, we apply the Scrabble score method to given names in Germany for the first time, to test the value of such scores to understanding the linguistic complexity of names in Germany. The corpus for these analyses contains 90.000 given names based on the SOEP data (1984–2016). We calculated the Scrabble score for all given names following the German Scrabble rules and analyzed different aspects. The results demonstrate that it is possible to also gain interesting insights about names in Germany using the Scrabble score method. For instance, we show that the frequencies of letters in given names differs (more vowels, more like vowels sounding consonants) from that in the other lexicon of the German language. Likewise, the method reveals gender-specific characteristics of male and female names. In addition, high Scrabble scores indicate non-German given names.

These three aspects of our study on names from the SOEP data illustrate that the Scrabble score method is suitable and productive for different questions about the linguistic complexity of given names in Germany.

JEL-Classification: Z13, C19

Keywords: Scrabble, given names, first names, onomastics, SOEP

Zusammenfassung

In US-amerikanischen Studien verschiedener Disziplinen hat sich zur Analyse der sprachlichen Komplexität von Vornamen seit einigen Jahren die so genannte Scrabble-Score-Analyse etabliert. Wie im bekannten Scrabble-Spiel Wörtern nach bestimmten Regeln Punktwerte zugeordnet werden, wird bei dieser Methode die Punktzahl einzelner Rufnamen ermittelt. Die mit ihrem jeweiligen Scrabblewert versehenen Rufnamen können anschließend auf flexible und quantitative Weise miteinander verglichen und ausgewertet werden.

Vorliegende Studie wendet die Scrabble-Score-Methode erstmals auf Rufnamen aus Deutschland an, um zu testen, welche Erkenntnisse über die sprachliche Komplexität von in Deutschland erhobenen Rufnamen möglich sind. Das zu Grunde liegende Rufnamenkorporus setzt sich aus rund 90.000 Rufnamen aus dem SOEP-Datensatz (1984-2016) zusammen. Für jeden einzelnen Namen ist nach den Scrableregeln ein Punktwert errechnet und Analysen zu verschiedenen Aspekten durchgeführt worden. Die Ergebnisse zeigen, dass mittels der Scrabble-Score-Methode auch im Deutschen interessante Erkenntnisse über Rufnamen gewonnen werden können. So lässt sich z.B. ermitteln, dass die Buchstabenverteilung im Deutschen für Rufnamen eine andere Gewichtung (mehr Vokale und mehr vokalisch klingende Konsonanten) aufweist als im übrigen deutschen Wortschatz. Ebenso werden geschlechtsspezifische Charakteristika von Frauen- und Männernamen deutlich. Darüber hinaus geben hohe Scrabblewerte Hinweise auf Rufnamen nicht-deutscher Herkunft.

Die Scrabble-Score-Methode zeigt damit beispielhaft anhand der drei hier ausgewählten Studienaspekten, dass sie für verschiedene Fragestellungen hinsichtlich der sprachlichen Komplexität der Rufnamengebung in Deutschland produktiv eingesetzt werden kann.

JEL-Classification: Z13, C19

Schlüsselwörter: Scrabble, Rufnamen, Vornamen, Onomastik, SOEP

1. Die Scrabble-Score-Methode in amerikanischen Namenstudien

Die Scrabble-Score-Methode ist eine völlig neuartige Methode zur Messung der sprachlichen Komplexität eines Namens, die 2010 erstmals von AURA/HESS in einer amerikanischen Namenstudie eingesetzt wurde. Die Autoren ermitteln in ihrer Untersuchung den sogenannten Scrabblewert der erhobenen Namen und vergleichen die Werte mittels Abstandsmessungen unter verschiedenen Gesichtspunkten miteinander (AURA/HESS 2010: 219). Der Scrabblewert ist eine jedem Namen zugeordnete Punktzahl, die sich ergibt, wenn man für die einzelnen Buchstaben festgelegte Punktwerte addiert. Der Punktwert richtet sich dabei nach der Gebrauchsfrequenz des jeweiligen Buchstabens im Lexikon einer bestimmten Sprache: häufige Buchstaben erhalten eine niedrige Punktzahl, seltener gebrauchte Buchstaben einen höheren Punktwert (vgl. Scrabble-Spielregeln)¹ in Scrabble Original Version 2012, Spielregeln online unter <http://www.scrabble.de/pdfs/Scrabble.pdf> . Durch dieses Punkteverfahren ist es möglich, ohne aufwendige linguistische Analysen den Fremdheitsindex von Namen im Vergleich zum appellativischen Lexikon der jeweiligen Sprache zu bestimmen.

Anhand der Scrabble-Score-Methode können AURA/HESS bspw. belegen, dass im amerikanischen Nameninventar anglo-amerikanische Rufnamen durchschnittlich einen niedrigeren Scrabblewert aufweisen als afro-amerikanische Rufnamen (AURA/HESS 2010: 222f.). Die Gründe dafür sind, dass letztere vermehrt im Englischen nicht so häufig verwendete Laute beinhalten und daher Buchstaben mit höheren Werten aufweisen sowie insgesamt auch eine längere Namenstruktur besitzen (vgl. dazu auch BALBACH 2018c: 320).

Die Scrabble-Score-Methode wurde in der amerikanischen Forschung von anderen Studien aufgegriffen und in kurzer Zeit als festes Instrumentarium etabliert. BIAVASCHI/GIULIETTI/SIDDIQUE (2017) gelingt es z. B. mit dieser Methode zu zeigen, dass Immigranten in den USA überwiegend dann ihren Namen ändern, wenn diese einen hohen Scrabblewert aufweisen und somit in ihrer sprachlichen Struktur stark vom amerikanischen Nameninventar abweichen. Gewählt werden von diesen Immigranten typische amerikanische Namen mit deutlich niedrigerem Scrabblewert als die Ursprungsnamen.

Zwei weiteren Studien² ist es mit Hilfe der Scrabble-Score-Methode gelungen, auf einen bedeutenden Umbruch in der amerikanischen Namenpraxis aufmerksam zu machen. Als Data Scientist wandte BERRY (2010) die Methode auf die von der US Social Security Administration herausgegebenen häufigsten amerikanischen Rufnamen zwischen 1880 und 2009 an. Seine Berechnungen zeigen, dass der durchschnittliche Scrabblewert der beliebtesten amerikanischen Rufnamen während dieses Zeitraums kontinuierlich angestiegen ist. Eine darauf aufbauende Studie der amerikanischen Namenforscherin WATTENBERG (2014) belegt, dass in den Jahren nach 2009 bis heute erstmals ein gegenteiliger Trend festzustellen ist und die Scrabblewerte wieder sinken. Als Grund dafür kann sie eine sich wandelnde amerikanischen Rufnamenpraxis ausmachen, die im Vergleich zu früher sowohl kürzeren als auch Namen mit vokaldominierter Lautstruktur den Vorzug gibt. Die Scrabble-Score-Methode ist daher nicht

¹ Scrabble-Spielregeln. In: Scrabble. Das Original. Version 2012. Auch online abrufbar unter: <http://www.scrabble.de/pdfs/Scrabble.pdf> (22.10.2018).

² Die Studien eines Informatikers und einer Namenforscherin werden hier aufgrund ihrer interessanten Ergebnisse genannt, auch wenn sie nicht wissenschaftlich, sondern nur in einem populär-wissenschaftlichen Blog publiziert wurden.

nur dazu geeignet, den Fremdheitsindex von Namen festzustellen. Vielmehr kann sie auch Wandel- und Entwicklungstendenzen von Namen innerhalb einer Sprache aufdecken.

Einen solchen phonologischen Wandel zu kürzeren und vokalreicheren Vornamen im Laufe der letzten Jahre hat für die deutsche Namengebung auch NÜBLING (2009) anhand umfangreicher phonologischer Analysen von Namen aus deutschlandweit ausgewählten Hitlisten festgestellt.

2. Anwendung der Scrabble-Score-Methode auf 90.000 Rufnamen des SOEP

2.1 Zum Namenkorpus und zur Datenaufbereitung

Zur exemplarischen Erprobung der Scrabble-Score-Methode an einem deutschen Namenkorpus ist aus dem SOEP-Datensatz (WAGNER ET AL. 2008, GOEBEL ET AL. 2018) aller Wellen bis einschließlich 2016 ein Korpus von rund 90.000 Rufnamen gebildet worden, die zwischen 1888 und 2016 an Personen vergeben wurden, die in Deutschland gelebt haben bzw. immer noch leben und an den Haushaltserhebungen im Rahmen der SOEP-Haushaltssurveys teilgenommen haben. Die 90.000 Rufnamen sind zufällig aus den rund 125.000 Namen umfassenden Wellen von 1984-2016³ ausgewählt worden, um ein etwas kleineres Versuchskorpus zu erstellen, das dennoch Rufnamen aus allen Wellen umfasst.

Diese 90.000 Rufnamen sind in einem nächsten Schritt einzeln daraufhin überprüft worden, ob es sich tatsächlich um Rufnamen handelt oder ob Kürzel (z. B. M., J.), Platzhalter (z. B. N.N., nn.) oder Familienverhältnisse (z. B. Mutter, Vater statt des Namens) verwendet wurden. Auch offensichtliche Falschschreibungen (z. B. Gnter statt Günter, Bnjamin statt Benjamin) von Namen wurden aus dem Korpus entfernt. Nach diesem Cleaning-Schritt umfasste das Korpus noch 89.943 Rufnamen. Der Einfachheit halber wird im Folgenden von „rund 90.000“ Rufnamen gesprochen. Sämtliche Werte (Prozentzahlen und andere Angaben) sind jedoch immer mit der exakten Rufnamenzahl berechnet worden.

Beim „Sozio-oekonomischen Panel (SOEP)“ handelt es sich um eine seit 1984 jährlich durchgeführte, repräsentative Haushaltserhebung unter dem Dach der Leibniz-Gemeinschaft. Im Rahmen dieser SOEP-Haushaltsbefragungen werden die Rufnamen sowie umfangreiche Hintergrundinformationen der Teilnehmer erhoben (vgl. HUSCHKA 2017). Diese Informationen betreffen u.a. Geburtsjahr, -ort, Geschlecht, Nationalität, Religion, Einkommen/Vermögen, Bildungsgrad und Beruf der teilnehmenden Personen.

Da die Rufnamen nur unter strengen datenschutzrechtlichen Richtlinien mit den Hintergrundinformationen zusammengespielt und für die Forschung zur Verfügung gestellt werden können, wurde im Rahmen dieses Tests der Scrabble-Score-Methode für die deutsche Namengebung erst einmal auf das Zusammenführen der Rufnamen mit ihren Hintergrundinformationen verzichtet. Das vorliegende Namenkorpus setzt sich daher aus rund 90.000 reinen Rufnamen ohne die dazugehörigen Hintergrundinformationen zusammen. Differenzierungen der Rufnamen nach z. B. Bildungsgrad oder Religion sind im Rahmen dieser Studie somit nicht möglich, für die methodische Erprobung aber auch nicht nötig.

Beachtet werden sollte für das folgende, dass es sich um ein Korpus von Rufnamen aus Deutschland handelt. Das Korpus besteht somit nicht allein aus Rufnamen von Deutschen (sie bilden zwar den Hauptteil), sondern aus Rufnamen von Personen verschiedener Nationalitäten, die in Deutschland

³ Es handelt sich um insgesamt 33 Wellen von 1984 bis einschließlich 2016, Datenversion v3x.x.

leben und in repräsentativer Auswahl an der SOEP-Haushaltserhebung teilgenommen haben (vgl. WAGNER ET AL. 2008, GOEBEL ET AL. 2018).

2.2. Zum methodischen Vorgehen

In Orientierung an AURA/HESS (2010) sind für jeden einzelnen Namen des vorliegenden Korpus die jeweiligen Scrabblewerte errechnet worden.⁴ Dabei ließen sich aus den rund 90.000 Rufnamen (token)⁵ 15.339 verschiedene Rufnamen (types) identifizieren, denen nach deutschem Scrabble-Punktesystem (vgl. Tabelle 1) die jeweiligen Werte zugeteilt wurden. Jeder Name wurde dabei so berücksichtigt, wie er im SOEP vorliegt: ist ein Doppelname angegeben, wurde auch ein Scrabblewert für den Doppelnamen errechnet. Schreibweisen mit bzw. ohne Bindestrich führen dabei zum gleichen Scrabblewert, da ein Bindestrich keinen Buchstabenwert hat und somit mit 0 Punkten bewertet wird. Gleiches gilt für das Leerzeichen zwischen zwei Namen, es erhält ebenfalls 0 Punkte bei der Berechnung. Entsprechend dem deutschen Scrabble-Punktesystem erhalten bspw. die Namen

Anna: 4 Scrabblepunkte,

Anton: 6 Scrabblepunkte,

Katharina: 13 Scrabblepunkte,

Hans Jürgen/Hans-Jürgen: 22 Scrabblepunkte und

Annemarie/Anne-Marie: 11 Scrabblepunkte.

Buchstabe	Punktewert
A, D, E, I, N, R, S, T, U	1
G, H, L, O	2
B, M, W, Z	3
C, F, K, P	4
Ä, J, Ü, V	6
Ö, X	8
Q, Y	10

Tab. 1: Deutsches Scrabble-Punktesystem
(aus Scrabble Original Version 2012)

⁴ Für die technische Unterstützung danke ich MICHAEL TIEKE und seinem Team von der IVV Geisteswissenschaften der Westfälischen Wilhelms-Universität Münster. SEBASTIAN BALBACH von der Universitätsklinik Münster danke ich für seine Unterstützung bei der Programmierung der computergestützten Verwaltung und Auswertung der rund 90.000 Vornamen.

⁵ In der quantitativen Linguistik wird zur Feststellung des Komplexitätsgrades eines Korpus (z. B. eines Textes, des Wortschatzes oder wie hier einer größeren Sammlung von Rufnamen) zwischen types und token unterschieden. Token sind jede einzelne vorkommende Wortform bzw. hier jeder einzelne vorkommende Rufname. Als types bezeichnet man die verschiedenen Wortformen bzw. hier jeden unterschiedlich vorkommenden Rufnamen. Bsp: In einer Schulklasse heißen fünf Mädchen Lilly und drei Mädchen Emma, eins heißt Marie und eins Sophie. Es gibt also zehn Mädchennamen in der Klasse (= token), aber nur vier verschiedene Namen (= types) (vgl. NÜBLING 2015: 22).

2.3 Zu den Ergebnissen

2.3.1 Häufige Scrabblewerte und Buchstabenfrequenz

Die meisten der 90.000 Rufnamen weisen einen Scrabblewert von 8 Punkten auf (N= 10.407 Namen, das sind 11,6% des Korpus) (vgl. Abbildung 1). Die häufigsten fünf Buchstaben in diesen Namen sind A (15,9%), E (10,4%), I (9,5%), N (9,0%) und R (8,1%), sie bilden zusammen über 50% aller in diesen Namen vorkommenden Buchstaben. Drei der fünf häufigsten Buchstaben sind Vokale (vgl. Abbildung 2).

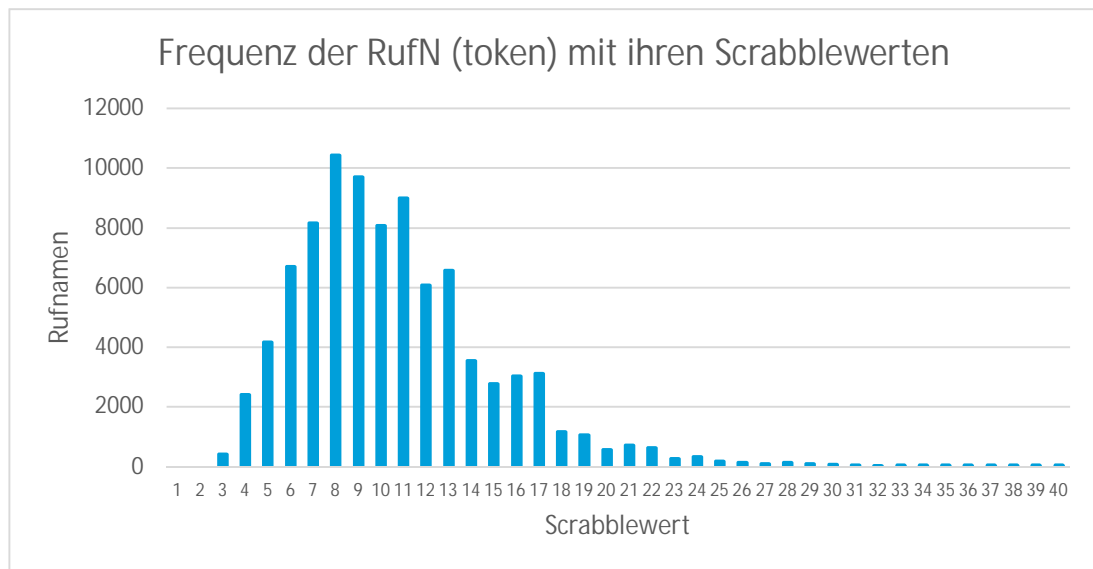


Abb. 1: Frequenz der Rufnamen (token) mit ihren Scrabblewerten

Daten: Sozio-oekonomisches Panel (SOEP), zufällige Namensauswahl von 90.000 Rufnamen aus den SOEP-Wellen 1984-2016 (Datenversion v33), eigene Berechnungen

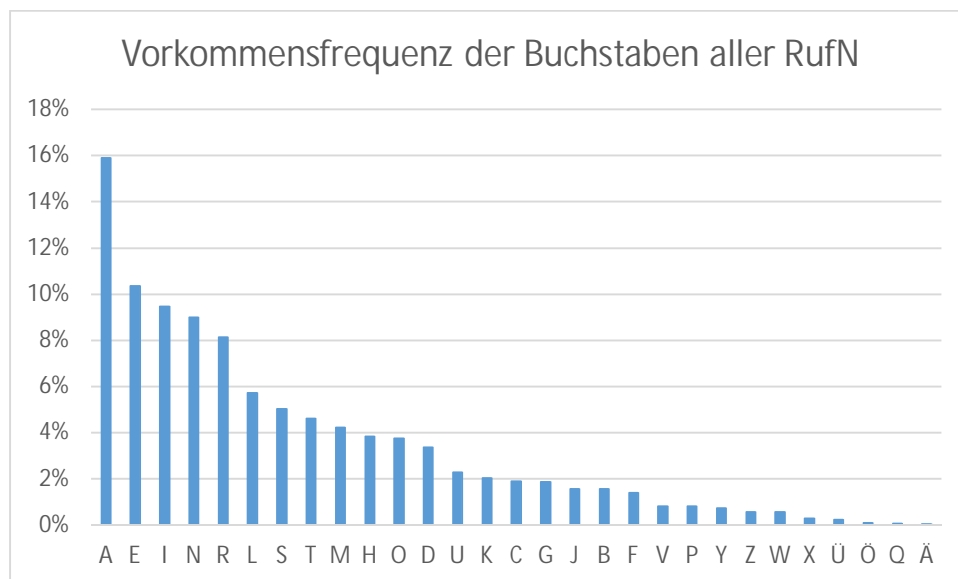


Abb. 2: Vorkommensfrequenz der Buchstaben aller Rufnamen (token)

Daten: Sozio-oekonomisches Panel (SOEP), zufällige Namensauswahl von 90.000 Rufnamen aus den SOEP-Wellen 1984-2016 (Datenversion v33), eigene Berechnungen

Zum Vergleich sei auf Abbildung 3 verwiesen, die die „Rangfolge der am häufigsten vorkommenden Buchstaben in allen Wortformen des Dudenkorpus“⁶ darbietet. Es wird deutlich, dass Vokale, vor allem der Vokal A, eine besondere Rolle innerhalb der Rufnamen spielen, da er sehr viel häufiger in Namen vorkommt als im übrigen deutschen Wortschatz (für den E an der Spitze liegt).⁷ Ebenso ist zu beobachten, dass Konsonanten wie L und M, die zu den so genannten Sonoranten gehören und damit „vokalisch klingendere Konsonanten“ sind, häufiger in Rufnamen verwendet werden als in anderen Wortformen. Eine phonologische Eigenschaft der Sonoranten ist es, dass sie weicher klingen als andere Konsonanten, da bei der Bildung der Sonoranten der Luftstrom kaum durch Zähne und Zunge behindert wird, sondern hauptsächlich die Stimmbänder vibrieren (vgl. NÜBLING 2009: 79). Ihre erhöhte Verwendung in Rufnamen zusammen mit dem ebenfalls vermehrten Einsatz von Vokalen lässt Rufnamen somit insgesamt weicher klingen.

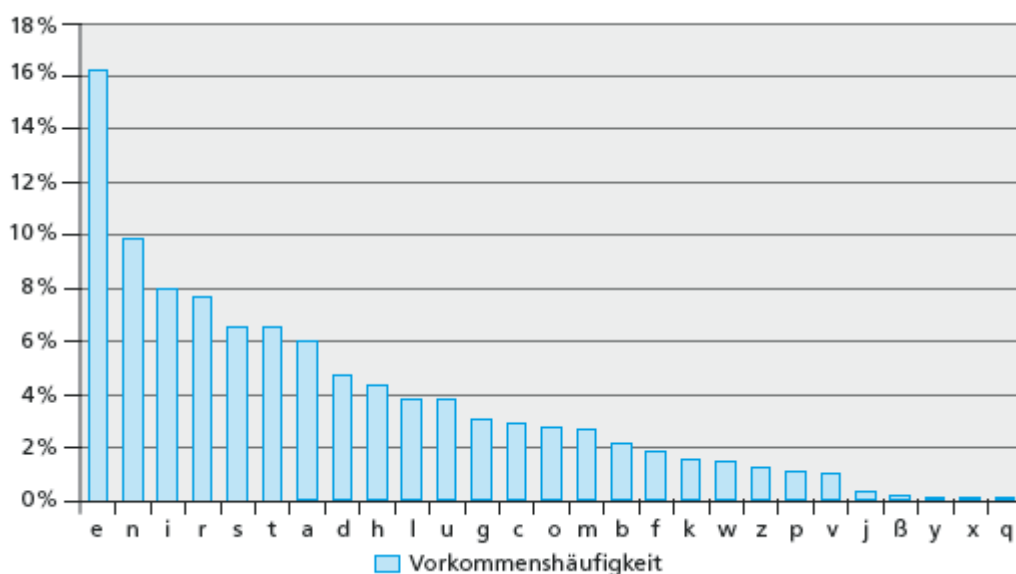


Abb. 3: Rangfolge der Buchstabenfrequenz in allen Wortformen des Dudenkorpus (DUDEN 2017: 148)

Die Analyse zeigt weiterhin, dass ein großer Teil der untersuchten Rufnamen ähnlich hohe Scrabblewerte aufweist: Namen mit Werten zwischen 7 und 11 Punkten bilden 50,3% aller 90.000 Rufnamen. Namen mit Werten zwischen 6 und 13 umfassen bereits 71,7% aller Rufnamen (vgl. Abbildung 1). Namen mit höheren Werten, v.a. mit Werten über 18 Punkten kommen nur selten vor. Insgesamt bilden die RufN zwischen 14 und 40 Scrabblepunkten nur 10,9% aller RufN.

⁶ DUDEN (Hg.) (2017): Die deutsche Rechtschreibung. 27., völlig neu bearbeitete und erweiterte Auflage. Berlin, S. 148. Abbildung 4 basiert auf über 4 Milliarden Wortformen. Die Abbildung kann auch im Duden online abgerufen werden: URL: <https://www.duden.de/sprachwissen/sprachratgeber/Die-hufigsten-Buchstaben-deutschen-Wortern> (15.10.2018).

⁷ Auch die unterschiedliche Auszählung der Umlaute kann die Differenzen der Vorkommensfrequenzen im Rufnamenkorpus und im Dudenkorpus nicht mindern. Zwar steigen durch die Auszählung der Umlaute als oe, ae, ue die Prozentzahlen der entsprechenden Vokale, doch ist immer noch deutlich zu sehen, dass die Vorkommensfrequenz bspw. des Vokals A im Rufnamenkorpus signifikant höher ist als im Dudenkorpus. Auch die übrige Vokal- und Konsonantenverteilung weist Unterschiede auf, die nicht durch die unterschiedliche Auszählung der Umlaute zu begründen ist.

Als erste Erkenntnis aus diesen Ergebnissen mittels der Scrabble-Score-Methode kann damit festgehalten werden, dass die meisten RufN im Deutschen eine ähnliche sprachliche Struktur aufweisen, da ihre Buchstabenwerte nah an einander liegen und nur wenige verschiedene Buchstaben benutzt werden. Vokale gehören dabei zu den bevorzugt verwendeten Buchstaben.

2.3.2. Geschlechtsspezifische Charakteristika der 100 häufigsten RufN im SOEP

In einem nächsten Schritt soll die Frage untersucht werden, ob sich geschlechtsspezifische Charakteristika der RufN mittels der Scrabble-Score-Methode feststellen lassen. Da sich die meisten Rufnamen im Deutschen eindeutig einem Geschlecht zuordnen lassen, sind für diese Analyse keine SOEP-Hintergrundinformationen nötig.

Aus dem Korpus der 90.000 RufN werden für diese Analyse die 100 häufigsten Namen nach Männer- und Frauennamen separiert und mit ihren Scrabblewerten zu einem neuen Teilkorpus komponiert. Es ergibt sich ein Korpus aus 55 MännerN und 45 FrauenN.

Die MännerN weisen mit durchschnittlich 10,8 Scrabblepunkten einen leicht höheren Scrabblewert auf als die FrauenN mit 9,0 Punkten. Zusammen mit der Tatsache, dass MännerN (5,9 Buchstaben) durchschnittlich etwas kürzer sind als FrauenN (6,1 Buchstaben), deutet dies auf eine leicht unterschiedliche sprachliche Struktur hin. Die genauere Analyse der Buchstabenhäufigkeit belegt dies: So sind die drei häufigsten Buchstaben in den analysierten FrauenN die Vokale A, I und E, sie bilden 42,0% aller in FrauenN verwendeten Buchstaben. Bei den MännerN stehen die Buchstaben A, R und E an der Häufigkeitsspitze. Sie stellen zusammen 33,4% aller in MännerN verwendeten Buchstaben. Nimmt man den viert häufigsten Buchstaben N mit 8,2% Vorkommensfrequenz hinzu, decken bei den MännerN zwei Vokale und zwei Konsonanten eine ähnlich hohe Buchstabenhäufigkeit ab wie innerhalb der FrauenN allein die drei Vokale A, I und E. Zudem ist festzustellen, dass unter den häufigsten FrauenN auch nur 21 verschiedene Buchstaben vorkommen, von denen der Vokal A hochfrequent ist, die Buchstaben I, E, N und R zwischen 8% und 12% liegen, die übrigen Buchstaben aber sehr viel seltener gebraucht werden (vgl. Abbildung 4). In MännerN hingegen ist die Buchstabenvielfalt mit 26 verschiedene Buchstaben nicht nur größer, auch ihre Verteilung konzentriert sich nicht so stark wie bei den FrauenN auf einige wenige häufiger genutzte Buchstaben. Wie in Abbildung 5 zu sehen, sinkt die Diagrammkurve der Buchstabenfrequenz moderater als bei den FrauenN.

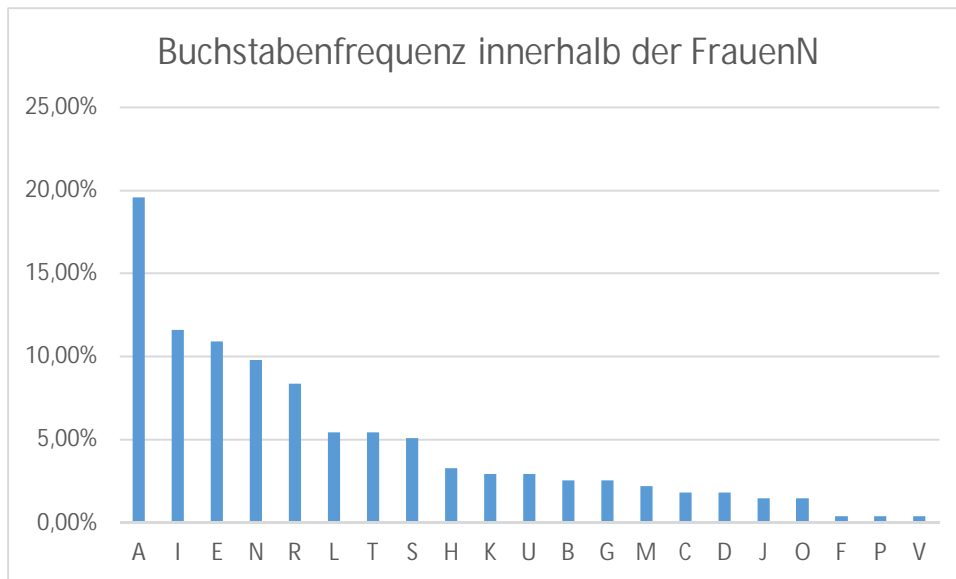


Abb. 4: Rangfolge der Buchstabenfrequenz innerhalb der Frauennamen (der 100 häufigsten Rufnamen). Buchstaben, die nicht abgebildet sind, kommen nicht vor (z. B. Ö, Ä)

Daten: Sozio-oekonomisches Panel (SOEP), zufällige Namensauswahl von 90.000 Rufnamen aus den SOEP-Wellen 1984-2016 (Datenversion v33), hier: nur die FrauenN aus den häufigsten 100 vergebenen RufN, eigene Berechnungen

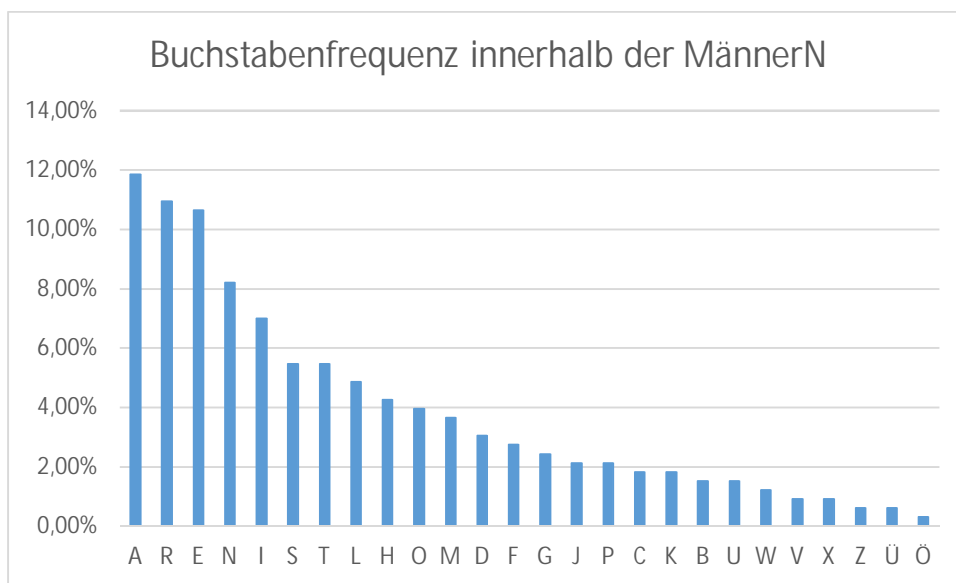


Abb. 5: Rangfolge der Buchstabenfrequenz innerhalb der Männernamen (der 100 häufigsten Rufnamen). Buchstaben, die nicht abgebildet sind, kommen nicht vor (z. B. Q)

Daten: Sozio-oekonomisches Panel (SOEP), zufällige Namensauswahl von 90.000 Rufnamen aus den SOEP-Wellen 1984-2016 (Datenversion v33), hier: nur die MännerN aus den häufigsten 100 vergebenen RufN, eigene Berechnungen

Die zweite Erkenntnis mittels der Scrabble-Score-Methode ist damit, dass sich Frauen- und MännerN in ihrer sprachlichen Struktur leicht unterscheiden und diese Unterschiede in der vokaldominanteren Struktur der FrauenN liegt. Bei der Errechnung der Scrabblewerte für die einzelnen Namen fällt zudem auf, dass sich höherwertige Buchstaben in MännerN oft im Auslaut des Namens befinden, sich aber Anlaut und Mitte der Namen meist aus Buchstaben mit ähnlichen Scrabblewerten wie in den FrauenN (hauptsächlich einwertige Buchstaben) zusammensetzen. Eine genauere Auslautanalyse der Männer- und FrauenN ergibt, dass die FrauenN in der untersuchten Stichprobe

1. nur sechs verschiedene Buchstaben als Endungen aufweisen (A, E, N, H, D, T),
2. die Vokale A und E bereits 84,4% aller Endungen bilden, die vier Konsonanten zusammen somit nur 15,6% der Endungen stellen.

Für die MännerN zeigt sich ein ganz anderes Ergebnis:

1. 15 verschiedene Buchstaben bilden hier die Endungen der häufigsten MännerN,
2. nur zwei davon sind Vokale (E und O), 13 hingegen Konsonanten (D, F, G, H, K, L, N, P, R, S, T, X, Z) und diese bilden zu 96,4% die Endungen der MännerN,
3. die häufigsten Endungen werden von den Buchstaben S und N gebildet (34,5%).

Als dritte Erkenntnis in Folge der Buchstaben-Scrabblewertbestimmung können damit Ergebnisse anderer Studien (z. B. GERHARDS 2003, NÜBLING 2009, OELKERS 2003) bestätigt werden, die ebenfalls feststellen konnten, dass Männer- und FrauenN im Deutschen eine starke Differenzierung der Namenendungen zeigen, die sich in einer eindeutigen Vokal- und Konsonantenpräferenz manifestiert. Während die bisherigen Studien, v.a. die Studien von NÜBLING (2009) und OELKERS (2003), dazu umfangreiche phonologische Analysen durchführen mussten,⁸ können die genannten geschlechtsspezifischen Charakteristika der 100 häufigsten deutschen RufN mittels der Scrabble-Score-Methode gleichermaßen einfach wie methodisch überzeugend nachgewiesen werden. Die Ergebnisse, die mit Hilfe der Scrabble-Score-Methode erzielt wurden, und mit den Erkenntnissen anderer Studien übereinstimmen, zeigen die Leistungsfähigkeit der hier angewendeten Methode und können somit als Validierung der Scrabble-Score-Methode gewertet werden.

2.3.3. RufN mit höheren Scrabblewerten als der Durchschnitt – Welche Besonderheiten weisen sie auf?

Wie unter Punkt 2.3.1. bereits erwähnt und in Abbildung 1 zu sehen, weisen die meisten der 90.000 RufN einen Scrabblewert von 8 Punkten auf (= 11,6% der RufN im Korpus). Als Durchschnittsscrabblewert aller RufN ergibt sich ein Wert von 10,6 Punkten. Rund ein Drittel der 90.000 RufN liegt allerdings mit seinen Scrabblewerten über dem Durchschnitt. Kann die Scrabblemethode hier Aufschluss darüber geben, wodurch sich diese RufN von denjenigen Namen unterscheiden, die näher am Scrabblewertdurchschnitt liegen?

Der Blick auf die RufN mit höheren Scrabblewerten zeigt, dass es sich vornehmlich um nicht-deutsche Namen handelt. Eine getrennte Analyse der RufN nach sprachlicher Herkunft der Namen (eingeteilt in deutsche und nicht-deutsche RufN),⁹ bestätigt, dass wesentlich mehr nicht-deutsche Rufnamen Werte

⁸ Die Studien von NÜBLING (2009) und OELKERS (2003) können aufgrund ihrer umfangreichen phonologischen Analysen aber auch tiefgehende Erkenntnisse über die lautlichen Differenzen von FrauenN und MännerN herausstellen, z. B. können sie Auskunft über differierende Silbenanzahl, Hauptakzente, Konsonanten-/Vokalverhältnisse und mehr geben.

⁹ Hierfür wurden alle 15.339 verschiedenen types der 90.000 Vornamen mit Hilfe folgender Werke in „deutsche“ und „nicht-deutsche“ Vornamen eingeteilt. Als deutsche Vornamen wurden diejenigen Namen definiert, die in der deutschen Namenpraxis frequent und seit Jahrzehnten etabliert sind, nicht ausschließlich die Namen, die sprachlich deutsche Wurzeln aufweisen. Dies ist aufgrund der christlich-jüdisch geprägten Namengebung gar nicht möglich, da die meisten deutschen Namen hebräische, griechische und lateinische Sprachwurzeln haben. Zur Einteilung verwendet wurden folgende Werke, die auch Aussagen über das Vorkommen der Namen im deutschen Nameninventar treffen: SEIBICKE (1996-2001), SCHIMMEL (1992), OLIVART (1993) sowie die Herkunftseinteilung, die in der Namenstudie von HUSCHKA (2017) verwendet wurde und der Autorin

zwischen 14 und 40 Scrabblepunkten aufweisen als deutsche Rufnamen, auch wenn der Hauptteil der nicht-deutschen Rufnamen ebenfalls in einem Scrabblewert-Spektrum von 6 bis 13 Punkten liegt. Allerdings bildet dieses Spektrum nur 63,1% aller nicht-deutschen RufN, während die höherwertigen RufN zwischen 14 und 40 Scrabblepunkten mit 26,4% mehr als doppelt so hoch liegen wie die deutschen RufN in diesem höheren Spektrum. Abbildung 6 mit der Darstellung der relativen Häufigkeitsverteilung der deutschen und nicht-deutschen RufN verdeutlicht die Verteilung. Die meisten deutschen RufN konzentrieren sich auf einige wenige Scrabblewerte, wie die Spitze des Balkendiagramms illustriert. Namen mit Scrabblewerten unter 6 Punkten und Namen mit Werten über 13 Punkten sind wesentlich seltener. Nicht-deutsche Namen hingegen verteilen sich breiter auf die verschiedenen Scrabblewerte, erreichen dadurch keine so hohen Spitzen wie die deutschen Namen und sinken nicht so stark im Bereiche der höheren Scrabblewerte ab. Im Gegenteil, die prozentuale Verteilung der nicht-deutschen Namen zeigt, dass sie im Vergleich zu den deutschen Namen durchgehend mehr Rufnamen mit Scrabblewerten zwischen 14 und 40 Punkten aufweisen (vgl. Abbildung 6).

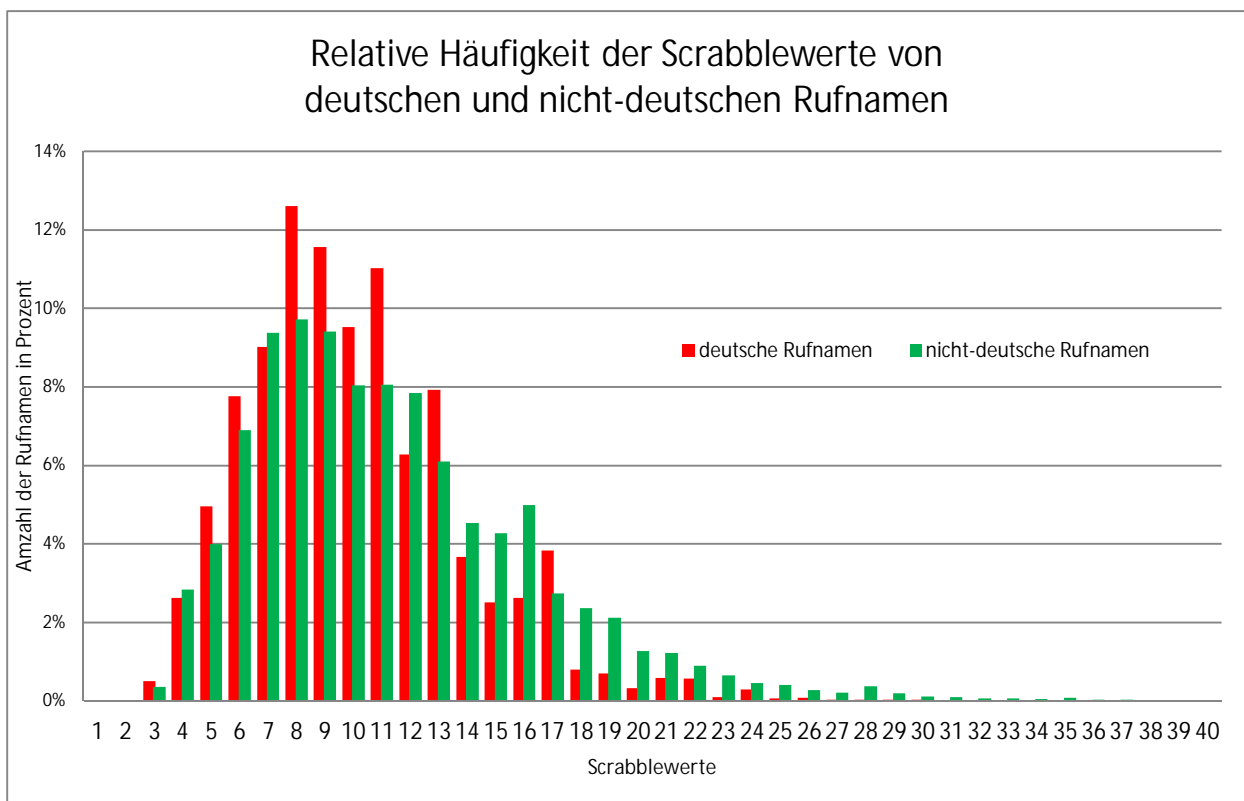


Abb. 6: Relative Häufigkeit der Scrabblewerte von deutschen und nicht-deutschen Rufnamen im Vergleich

Daten: Sozio-oekonomisches Panel (SOEP), zufällige Namensauswahl von 90.000 Rufnamen aus den SOEP-Wellen 1984-2016 (Datenversion v33), eigene Berechnungen

freundlicherweise zur Verfügung gestellt wurde. HUSCHKAS Herkunftseinteilung wurde von Frau Gabriele Rodriguez erstellt, die als Fachberaterin für Vornamen an der Namenberatungsstelle der Universität Leipzig tätig ist und als Expertin herangezogen wurde.

Für weitere Namenstudien kann das Zuspänspielen der Information „Nationalität“ aus dem SOEP-Datensatz vorgenommen werden, dies wurde für die vorliegende Methodenprüfung aufgrund der hohen Datenschutzbestimmungen noch nicht durchgeführt. Vgl. dazu auch LIEBAU ET AL. (2018), besonders S. 9f.

Die Ursache für die unterschiedlichen Scrabblewerte deutscher und nicht-deutscher RufN liegt nicht in einer differierenden Namenlänge. Die Durchschnittslänge deutscher RufN ist 6,2 Buchstaben, die der nicht-deutschen RufN ist mit 6,18 Buchstaben fast identisch. Korreliert man die Buchstabenlänge der Namen und mit ihrem Scrabblewert zeigt sich, dass ein längerer Rufname (= mehr Buchstaben) nicht unmittelbar zu höheren Scrabblewerten führt, sondern auch kürzere Namen hohe Scrabblewerte erreichen können. Es sind die einzelnen Laute, in der Scrabble-Score-Methode durch Buchstaben gemessen, die zu ganz unterschiedlichen Scrabblewerten führen. Das Korpus beinhaltet bspw. Rufnamen mit 5 Buchstaben, die Scrabblewerte zwischen 5 und 31 Punkten aufweisen. Ebenso variieren Rufnamen mit bspw. 15 Buchstaben zwischen Scrabblewerten zwischen 15 und 41 Punkten. Das bedeutet, dass die Länge des Rufnamens vor allem Einfluss auf die Mindestpunktzahl hat: Ein Rufname mit 10 Buchstaben verfügt mindestens über einen Scrabblewert von 10 Punkten, da jeder Buchstabe mindestens einen Punktwert von 1 erhält. Die Maximalhöhe des Scrabblewertes variiert allerdings stark: In vorliegendem Namenkorpus weisen dreibuchstabige Rufnamen Scrabblewerte zwischen 3 und 18 Punkten auf, sechsbuchstabile Rufnamen liegen zwischen 6 und 31 Punkten. Für die weitere Verteilung der Werte vgl. Abbildung 7, in der die Buchstabenlänge der Rufnamen mit ihren Scrabblewerten korreliert und graphisch dargestellt ist.

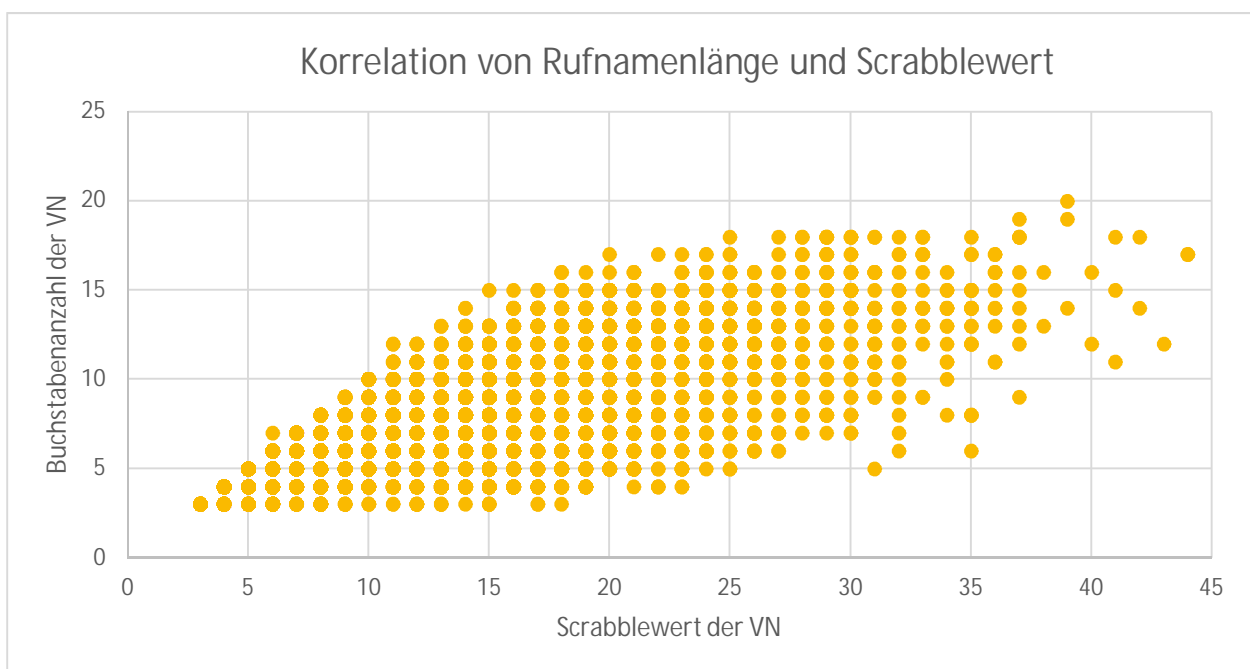


Abb. 7: Korrelation von Rufnamenlänge (Buchstabenanzahl) und Scrabblewert

Daten: Sozio-oekonomisches Panel (SOEP), zufällige Namensauswahl von 90.000 Rufnamen aus den SOEP-Wellen 1984-2016 (Datenversion v33), eigene Berechnungen

Die Analyse der Buchstabenhäufigkeit zeigt, dass die höheren Scrabblewerte der nicht-deutschen Namen durch die Verwendung von Buchstaben mit im Deutschen hohen Punktwerten zu Stande kommt und somit Laute bzw. Schreibweisen verwendet werden, die im Deutschen weniger häufig sind. So treten in den nicht-deutschen Namen bspw. vermehrt die Buchstaben W, Z, C, K, P, Q und Y auf, die drei und vier Scrabblepunkte erhalten bzw. Q und Y sogar je zehn Scrabblepunkte. Vor allem die häufigere Verwendung des Buchstabens Y in nicht-deutschen Namen führt auch in kurzen Namen zu hohen Scrabblewerten. Y ist in nicht-deutschen Namen mit 12,7% weitaus häufiger zu finden als in deutschen Namen (nur 2,9%).

Die Scrabble-Score-Methode leitet damit zu der vierten Erkenntnis, dass deutsche RufN aufgrund ihrer ähnlichen sprachlichen Komplexität auch ähnliche Scrabblewerte aufweisen. Davon abweichende, vor allem höhere Scrabblewerte deuten auf Namen nicht-deutschen Ursprungs hin, was durch Buchstabenhäufigkeitsanalysen belegt werden kann.

3. Fazit und Ausblick

Die Scrabble-Score-Methode liefert schon an einem reinen Rufnamenkörper (ohne weitere Hintergrundinformationen zu den Namensträgern) interessante Erkenntnisse. Breitere und tiefergehende Ergebnisse sind zudem möglich, wenn die RufN um die verschiedenen sozioökonomischen Hintergrundinformationen zu den einzelnen Namensträgern ergänzt werden. So wären z. B. Einblicke in die spezifische Namenwahl bestimmter Berufs- und Bildungsgruppen (Schichtenspezifik), in regionale Namenwahlen (Regionenspezifik), in Namenwahlen aufgrund bestimmter Religionszugehörigkeit und aufgrund verschiedener Migrationshintergründe möglich.

Für Nicht-Linguisten stellt die Scrabble-Score-Methode eine einfache Möglichkeit dar, Aussagen über die sprachliche Komplexität eines Namens zu machen und die Ergebnisse als Instrument für jeweils disziplinspezifische Fragestellungen zu nutzen (vgl. z. B. die amerikanischen Studien von AURA/HESS 2010, BERRY 2010, BIAVASCHI ET AL. 2017, WATTENBERG 2004, 2014). Aber auch Linguisten können die Methode nutzen, um Hypothesen zu prüfen und lohnende Ansatzpunkte für tiefergehende linguistische Analysen zu finden. Durch das Zusammenspielen der Rufnamen und der Geburtsjahrgänge der benannten Personen können bspw. nach Vorbild von BERRY (2010) und WATTENBERG (2014) Rufnamenentwicklungen und Wandelerscheinungen der Namenpraxis der rund 100 letzten Jahre aufgedeckt werden. Besonders Zäsur trüchtige Zeiträume könnten identifiziert und mittels weiterführender Analysen genauer untersucht werden. Studien zu diesen und anderen Aspekten sind in einem linguistisch-onomastischen Projekt zur Rufnamengebung des 20. und 21. Jahrhunderts auf Basis des SOEP-Rufnamenkörper und -Datensatzes geplant.

Literatur

Aura, Saku/Hess, Gregory D. (2010): What's in a Name? In: *Economic Inquiry* 48/1, 214–227.

Balbach, Anna-Maria (2018c): Von Agustin über Tom zu DaShawn – Zur Geschichte und Entwicklung so genannter ‚Black Names‘ in den USA. Teil 3: Von der Mitte des 20. Jahrhunderts bis heute In: *Beiträge zur Namenforschung* 53/3, 283–338.

Berry, Nick (2010): What's in a name? In: *DataGenetics*. URL: <http://datagenetics.com/blog/january12011/index.html> (19.10.2018).

Biovaschi, Costanza/Giulietti, Corrado/Siddique, Zahra (2017): The Economic Payoff of Name Americanization. In: *Journal of Labor Economics* 35/4, 1089–1116.

Duden (Hg.) (2017): *Die deutsche Rechtschreibung*. 27., völlig neu bearbeitete und erweiterte Auflage. Berlin.

Goebel, Jan/Grabka, M., Liebig, S., et al. (2018): The German Socio-Economic Panel (SOEP). In: *Jahrbücher für Nationalökonomie und Statistik*, 0(0), 1-16. Retrieved 15 Oct. 2018, from doi: 10.1515/jbnst-2018-0022.

Huschka, Denis (2017): *Naming in Germany in the 20th Century. A Sociological Study of Naming in Times of Social change, with a Focus on Statistical Problems in Empirical Onomastic Research*. Dissertation. Grahamstown/Südafrika.

- Liebau, Elisabeth/Humpert, Andreas/Schneiderheinze, Klaus (2018): Wie gut funktioniert das Onomastik-Verfahren? Ein Test am Beispiel des SOEP-Datensatzes. In: SOEPpapers on Multidisciplinary Panel Data Research 976, 1-28.
- Nübling, Damaris (2009): Von Horst und Helga zu Leon und Leonie: Werden die Rufnamen immer androgyner? In: Der Deutschunterricht 5, 77-83.
- Nübling, Damaris/Fahlbusch, Fabian/Heuser, Rita (2015): Namen. Eine Einführung in die Onomastik. 2., überarbeitete und erweiterte Auflage. Tübingen.
- Oelkers, Susanne (2003): Naming Gender. Empirische Untersuchungen zur phonologischen Struktur von Vornamen im Deutschen. Frankfurt.
- Olivart, Jose Maria Albaiges (1993): Diccionario de Nombres de Personas. Universidad De Barcelona. Publicaciones Y Ediciones. Barcelona.
- Schimmel, Annemarie (1992): Herr „Demirci“ heißt einfach „Schmidt“. Türkische Namen und ihre Bedeutung. Köln.
- Scrabble-Spielregeln. In: Scrabble. Das Original. Version 2012. Auch online abrufbar unter: <http://www.scrabble.de/pdfs/Scrabble.pdf> (22.10.2018).
- Seibicke, Wilfried (1996-2001): Historisches Deutsches Vornamenbuch. Band 1-5. Berlin.
- Sozio-oekonomisches Panel (SOEP), Daten für die Jahre 1984-2016, Version 33, SOEP, 2017, doi:10.5684/soep.v33.
- Wagner, Gert G./Göbel, Jan/Krause, Peter/Pischner, Rainer/Sieber, Ingo (2008): Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland - Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender) In: AStA Wirtschafts- und Sozialstatistisches Archiv, Jg. 2, No. 4, 301-328.
- Wattenberg, Laura (2004): Name Inflation, Scrabble Score. In: Babynamewizard. URL: <http://www.babynamewizard.com/archives/2004/9/name-inflation-scrabble-style> (19.10.2018)
- Wattenberg, Laura (2014): How do Baby Names Score at Scrabble? In: Babynamewizard. URL: <http://www.babynamewizard.com/archives/2014/9/how-do-baby-names-score-at-scrabble-special-10th-anniversary-edition> (19.10.2018).