

Für beide Varianten können – bei entsprechender Begründung und einer Darstellung der Wirtschaftlichkeit der jeweiligen Lösung – Fördermittel bei der DFG beantragt werden. Da durch eine Förderung mit Drittmitteln außerplanmäßige Ressourcen erschlossen werden können, bietet sich gerade bei kleineren Archiven die Vergabe an Dienstleister an. Über den Internetauftritt der Archivschule Marburg können auch weiterhin qualifizierte Dienstleister ermittelt werden.

Wichtig ist für einen erfolgreichen Projektabschluss auch, das Projektende genauso gut vorzubereiten, wie die Antragstellung oder die eigentliche Retrokonversionsarbeit. Gemeint ist die Datenlieferung an ein entsprechendes Online-Portal. Die Rückmeldungen derjenigen, die bislang Retrokonversionsprojekte durchgeführt haben, geht dahin, dass dieser Projektschritt – bisweilen mit Unterstützung von Dritten – auch für digital unerfahrene Kolleginnen und Kollegen bewältigbar ist. Allerdings sollten hier die technischen Details nicht erst bedacht werden, wenn alle Retrokonversionsprodukte vorliegen. Frühzeitige Kontaktaufnahme mit den Portalbetreibern oder Probeexporte aus den Softwareanwendungen helfen später bei der zügigen Online-

Stellung. In mehreren Fällen konnte durch eine Antragsstellung den jeweiligen Trägerverwaltungen der Wertarchivischer Fachsoftware plausibel gemacht werden.

Bei der Frage der Online-Präsentation der Findmitteldaten hat in den letzten sechs Jahren eine Konzentration auf etwa ein Duzend Archivportale stattgefunden. Die Implementierung des Archivportals-D wird hier einen weiteren Mehrwert für die Forschung schaffen. Für die Archive lag der Mehrwert nach Rückmeldung von Projektteilnehmern neben der verstärkten Präsenz in der Öffentlichkeit und einer stärkeren Nutzung der Bestände auch in einer engeren kollegialen Vernetzung. Dieser fachliche Austausch und die so genutzten Netzwerke können nun dazu beitragen, die archivische Fachaufgabe Retrokonversion auch ohne die Unterstützung durch eine Koordinierungsstelle anzugehen. Die Deutsche Forschungsgemeinschaft stellt auch weiterhin erhebliche Mittel für diese Aufgaben zur Verfügung. Aktuelle Informationen zum Antragsprozess können über den Webauftritt der DFG oder direkt bei Frau Dr. Franziska Regner (Franziska.Regner@dfg.de Telefon: +49 228 885-2094) abgerufen werden. ■

Claudius Kienzle, Stuttgart

„GIBT ES GUTE BUCHSTABEN?“

UNTERSUCHUNG ZUR EIGNUNG BESTIMMTER BUCHSTABEN FÜR KLUMPENSTICHPROBEN

AUSGANGSLAGE

2012 beschloss das Staatsarchiv Zürich, seine Stichprobenverfahren bei der Bewertung von massenhaft gleichförmigen Einzelfallakten zu überprüfen und neue Verfahren auszuprobieren: Einerseits wurde der von uns teilweise gewählte Buchstabe B für Klumpenstichproben aus alphabetisch abgelegten Unterlagen in Frage gestellt, andererseits wollten wir erstmals eine Zufallsstichprobe nach statistischen Grundsätzen ziehen.

Überall dort, wo weder die Vollarchivierung eines Bestandes noch dessen vollständige Kassation in Frage kamen, wendete das Staatsarchiv Zürich bisher folgende Auswahlmethoden an:

- Archivierung von Mustern;
- inhaltliche Auswahl von Dossiers nach bestimmten Kriterien;
- systematische Stichprobe (jeder x-te Fall);
- Klumpenstichprobe (bei chronologischen Ablagen oft jeder 5. Jahrgang, bei alphabetischen Ablagen der Buchstabe B).

Als theoretische Grundlage für die Untersuchung der Tauglichkeit des Buchstabens B sowie die Ziehung einer reinen Zufallsstichprobe¹ dienen uns das Statistik-Lehrbuch von Peter Bohley², die Fallstudie zu Stichprobenverfahren von Matthias Buchholz³ und das Papier der Arbeitsgruppe Bewertung des VSA (Verein Schweizerischer Archivarinnen und Archivare) zum Thema „Stichprobenziehung/Sampling“⁴.

Für kleinere Grundgesamtheiten von Unterlagen sind Zufallsstichproben nach statistischen Grundsätzen nicht geeignet. Zudem ist deren Ziehung relativ aufwändig. Für zahlreiche alphabetisch geordnete Aktengruppen nimmt das Staatsarchiv Zürich deshalb eine Klumpenstichprobe nach Anfangsbuchstaben vor – und zwar werden alle Akten, deren Ordnungsbegriff mit dem Buchstaben B beginnt, übernommen. Diese Auswahl ist nicht zuletzt aufgrund ihrer einfachen Handhabung interessant. So kann sie auch von den abliefernden Stellen selber durchgeführt werden. Obwohl die Buchstabenauswahl nicht repräsentativ ist und es insofern keine

„guten“ oder „schlechten“ Buchstaben gibt, zeigt sich rasch, dass sich mit bestimmten Buchstaben eine bessere Annäherung an Repräsentativität erreichen lässt als mit anderen. Deshalb wollten wir genauer untersuchen, welche Buchstaben sich am besten eignen und was das für den bisher angewendeten Buchstaben B heißt.⁵

VORGEHEN

Da die meisten Buchstabenauswahlen bei nach Namen geordneten Akten getroffen werden, haben wir uns mit dieser Art von Stichproben und mit der Verteilung von Anfangsbuchstaben näher befasst. Ausgehend von der bei Buchholz⁶ besprochenen Thematik der Buchstabenauswahl mit ihren Vor- und Nachteilen, wollten wir bezüglich Anfangsbuchstaben des Nachnamens zwei Fragen untersuchen:

1. Wie groß ist der Anteil der einzelnen Anfangsbuchstaben am Total aller Nachnamen?
2. Wie verteilen sich die Anfangsbuchstaben über verschiedene Nationalitäten?

Der erste Punkt ist für eine sinnvolle Stichprobengröße wichtig, der zweite soll zeigen, welche Buchstaben in welchen Nationalitäten (Sprachen) überhaupt vorkommen und bei welchen die Verteilung auf die Nationalitäten am angemessensten ist. Unter Umständen wären neben der Nationalität weitere Merkmale (etwa die Religionszugehörigkeit oder die Zugehörigkeit zu einer sozialen Schicht) denkbar, die eine bestimmte Korrelation mit den Anfangsbuchstaben von Nachnamen haben. Wir haben uns jedoch auf die Nationalität beschränkt.

Als Untersuchungsraum wählten wir zuerst unseren Archivsprengel, den Kanton Zürich. Weil beim Statistischen Amt des Kantons Zürich jedoch keine Daten zu Familiennamen vorhanden sind, wurde als Datenbasis die Wohnbevölkerung in der Stadt Zürich am Ende des Jahres 2011 verwendet. Statistik Stadt Zürich lieferte uns freundlicherweise eine Kreuztabelle nach den Anfangsbuchstaben des Nachnamens und ausgewählten Nationalitäten. In die Auswahl aufgenommen wurden die acht über die letzten zwanzig Jahre in der Stadt Zürich am stärksten vertretenen Nationalitäten. Die restlichen Nationalitäten wurden unter „Übrige“ zusammengefasst. Dies ergab folgende Liste (nach Häufigkeit geordnet): Schweiz, Deutschland, Italien, Portugal, Serbien/Montenegro/Kosovo, Spanien, Türkei, Sri Lanka, Übrige.

AUSWERTUNG

Auf der Basis der gelieferten absoluten Zahlen nahmen wir folgende Auswertungen vor:

1. Anteil eines Buchstabens am Total einer Nationalität (Beispiel: 10,64 % aller Schweizer/innen in der Stadt Zürich haben einen Nachnamen, der mit B beginnt);
2. Anteil einer Nationalität am Total eines Buchstabens (Beispiel: 75,58 % aller Einwohner/innen in der Stadt Zürich, deren Nachname mit B beginnt, sind Schweizer/innen);
3. Anteil einer Nationalität an der Gesamtbevölkerung (Beispiel: 68,98 % aller Einwohner/innen der Stadt Zürich sind Schweizer/innen);
4. Verhältnis von Anteil einer Nationalität am Total eines Buchstabens zu mittlerem Anteil dieser Nationalität an der Gesamtbevölkerung (Beispiel: Verhältnis von 2 zu 3).⁷

Aufgrund der ersten Auswertung lässt sich etwas über die Grösse der Stichprobe sagen, die mit der Entnahme eines Buchstabens entsteht. So zieht man beispielsweise mit dem Buchstaben B

eine Stichprobe von knapp zehn Prozent, mit dem Buchstaben R hingegen eine von gut fünf Prozent.

	Anteil an Total [%]		Anteil an Total [%]		Anteil an Total [%]
A	4,29	J	1,78	S	12,71
B	9,71	K	6,62	T	3,04
C	3,57	L	4,05	U	0,53
D	4,18	M	8,34	V	2,44
E	2,16	N	2,08	W	4,14
F	4,05	O	1,36	X	0,05
G	5,74	P	3,88	Y	0,36
H	6,33	Q	0,11	Z	2,12
I	1,11	R	5,24	Total	100,00

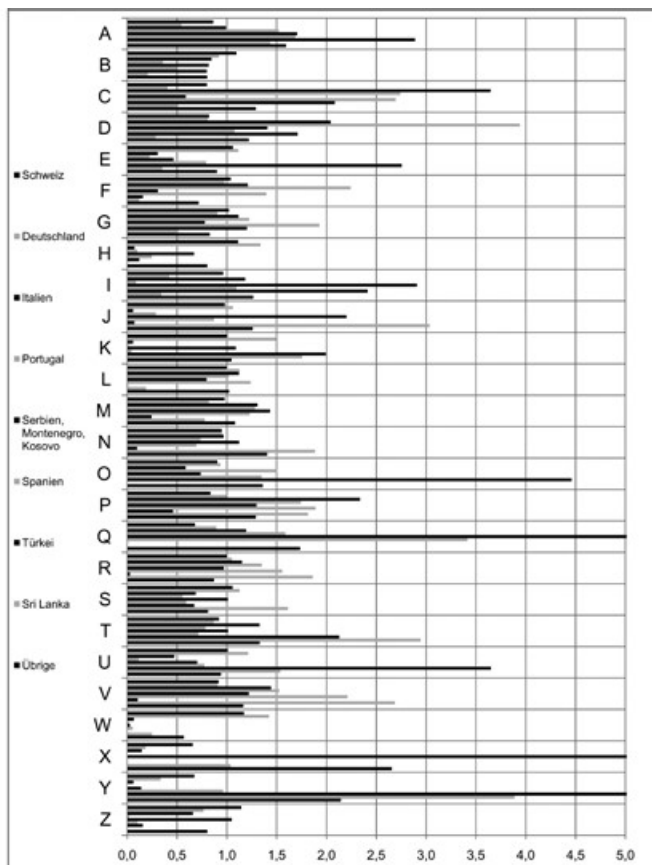
Tab. 1: Anteil von Einwohner/innen, deren Nachname mit einem bestimmten Buchstaben beginnt, an der Gesamtheit aller Einwohner/innen

Das in der vierten Auswertung dargestellte Verhältnis (vgl. Abb. 1) sagt aus, wie häufig ein Anfangsbuchstabe in den Nachnamen einer bestimmten Nationalität vorkommt, wenn man zusätzlich den Anteil dieser Nationalität an der Gesamtbevölkerung mitberücksichtigt. Wie nicht anders zu erwarten, zeigen sich große Unterschiede unter den Buchstaben: Einige kommen bei bestimmten Nationalitäten gar nicht oder sehr selten vor, andere wiederum sind deutlich übervertreten. Diese Extremwerte sind mit linguistischen Besonderheiten der jeweiligen Sprachen, die bei den erfassten Nationalitäten am häufigsten gesprochen werden, zu begründen. So kommt etwa der Anfangsbuchstabe X bei Portugiesen, Spanierinnen und Türken sowie der Anfangsbuchstabe O bei Personen aus Sri Lanka gar nicht vor, während der Buchstabe H praktisch nur bei Schweizerinnen, Deutschen und Personen aus Serbien, Montenegro und Kosovo vertreten ist. Sehr hohe Werte ergeben sich wiederum bei bestimmten Nationalitäten für in der Schweiz ansonsten eher exotische Anfangsbuchstaben wie X (Serbien, Montenegro, Kosovo = 906) oder Y (Türkei = 1969). Alle Buchstaben mit solchen Extremwerten fallen für eine Buchstabenstichprobe außer Betracht, es sei denn, es könnten mehrere Buchstaben komplementär miteinander kombiniert werden. Darauf kommen wir noch zurück.

⁵ Die Zufallsstichprobe wurde am Beispiel von 15 Aktenjahrgängen der kantonalen Baurekurskommissionen mit insgesamt mehr als 21.000 Dossiers durchgeführt. Auf die konkrete Durchführung wird hier nicht näher eingegangen, da sie dem in der einschlägigen Archivliteratur empfohlenen Vorgehen entspricht. Wer sich dafür interessiert, kann sich gerne an die Abteilung Überlieferungsbildung des Staatsarchivs Zürich wenden (thomas.neukom@jj.zh.ch).

⁷ Peter Bohley: Statistik. Einführendes Lehrbuch für Wirtschafts- und Sozialwissenschaftler. 7. Aufl. München 2000, Kapitel 16, S. 497–515.

Auf der Suche nach Buchstaben, die bei allen untersuchten Nationalitäten vorkommen und die dabei eine ziemlich ausgeglichene Verteilung zeigen, fällt das Augenmerk vor allem auf folgende vier: B, G, M und S. Allerdings muss auch bei diesen vier Buchstaben betont werden, dass die Unterschiede zwischen den Nationalitäten innerhalb eines Buchstabens immer noch beträchtlich sind: Die höchsten Werte sind um einem Faktor 3 (S), 4 (G) oder 5 (B, M) größer als die niedrigsten.



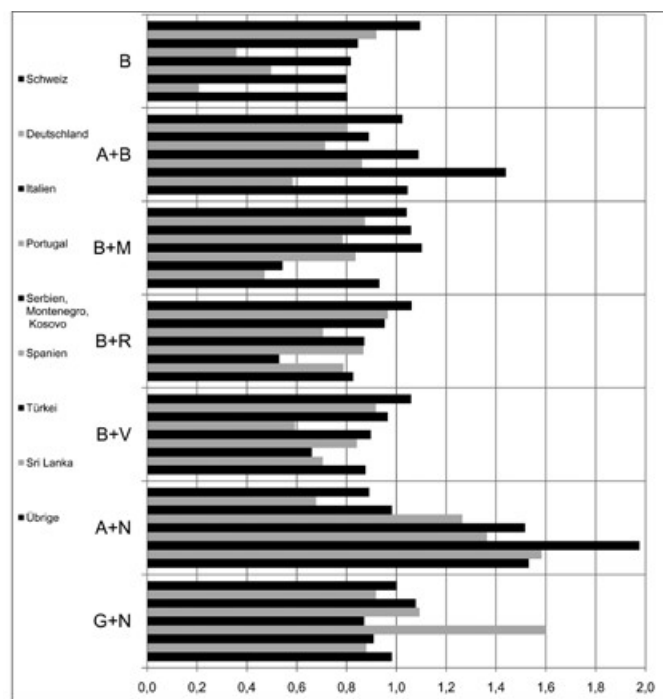
Verhältnis von Anteil einer Nationalität am Total eines Buchstabens zu mittlerem Anteil dieser Nationalität an der Gesamtbevölkerung
Horizontale Achse bei 5 gekappt (Werte: Serbien, Montenegro, Kosovo Q = 6.05 und X = 9.06; Türkei Y = 19.69)

Da die Daten „nur“ aus dem Gebiet der Stadt Zürich stammen, wir uns aber eigentlich für den ganzen Kanton Zürich interessierten, stellten wir zusätzlich folgende Berechnung an: Unter der Annahme, dass die Verteilung der Anfangsbuchstaben bei einer Nationalität in der Stadt und im ganzen Kanton gleich ist, wurden die oben unter Punkt 2 erwähnten Zahlen (Anteil einer Nationalität am Total eines Buchstabens) durch die kantonalen Anteile einer Nationalität an der Gesamtbevölkerung geteilt, da sich letztere teilweise markant von den Verhältnissen in der Stadt Zürich unterscheiden. Dies führte zu einer Variante des oben unter Punkt 4 erwähnten Verhältnisses (Verhältnis von Anteil einer Nationalität am Total eines Buchstabens zu mittlerem Anteil dieser Nationalität an der Gesamtbevölkerung). Auch wenn sich bei dieser Berechnung einige Extremwerte deutlich veränderten, so blieb das Gesamtbild doch gleich. Vor allem zeigte sich auch hier wieder, dass die Verteilung der Nationalitäten bei den Buchstaben B, G, M und S am ausgeglicheneren ist.

Für die bisher im Staatsarchiv Zürich angewendete Auswahl des Buchstabens B kann bezüglich der zwei eingangs gestellten Fragen also gefolgert werden:

1. Der Anteil der Personen in der Stadt Zürich, deren Nachname mit B beginnt, beträgt 9,71 %. Das bedeutet, dass eine B-Stichprobe bei der Aktenauswahl in der Regel etwa einer Auswahl von 10 % der Gesamtmenge entspricht.
2. Das Verhältnis von Anteil einer Nationalität am Total eines Buchstabens zum mittleren Anteil dieser Nationalität an der Gesamtbevölkerung ist beim Buchstaben B für die gebildeten Gruppen relativ ausgeglichen, und zwar für beide Rechnungsvarianten (Stadt Zürich und Kanton Zürich).

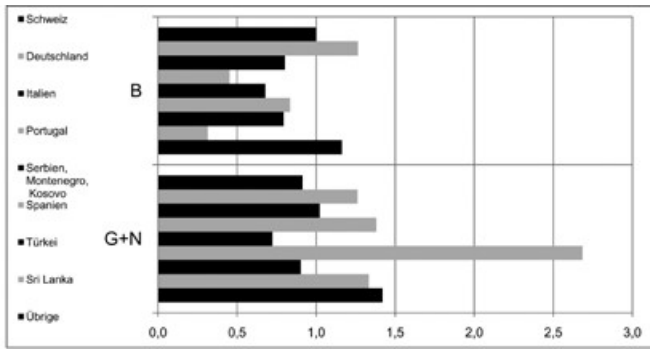
Zur Abklärung der Frage, ob sich Ungleichheiten bei einzelnen Buchstaben durch eine Kombination mehrerer Buchstaben noch etwas ausgleichen ließen, stellten wir einige Zusatzberechnungen für bestimmte Buchstabenpaare an.



Verhältnis von Anteil einer Nationalität am Total eines Buchstabens zum mittleren Anteil dieser Nationalität an der Gesamtbevölkerung, Buchstabe B sowie sechs Kombinationen von zwei Buchstaben.

Dabei zeigt sich, dass geschickt gewählte Buchstabenpaare ausgeglichener Zahlen liefern als einzelne Buchstaben. So wäre beispielsweise die Kombination G+N für die Stadt Zürich sehr interessant. Mit Ausnahme des Wertes für Spanien zeigt sich hier eine sehr hohe Ausgeglichenheit. Allerdings relativiert sich diese Aussage bereits wieder, wenn man die kantonalen Vergleichswerte betrachtet.

Zudem muss noch in Betracht gezogen werden, welche Mengenanteile man mit einer Kombination von Buchstaben aus der Gesamtheit herauszieht. Mit dem Buchstaben B erhält man ein Sample von knapp zehn Prozent. Will man eine Stichprobe also nicht grösser als zehn Prozent werden lassen, so bietet sich der Buchstabe B nicht unbedingt für eine Kombination mit weiteren Buchstaben an. Hingegen wären A+N mit gut sechs Prozent oder G+N mit knapp acht Prozent gut geeignet.



Vergleichszahlen zu Abb. 2 mit kantonalen Zahlen für den mittleren Anteil einer Nationalität an der Gesamtbevölkerung, Buchstabe B sowie Kombination von G

Was sich zusätzlich zeigte: Auch wenn eine Buchstabenauswahl kein repräsentatives Sample darstellt, lässt sie eher einen Rückschluss auf die Grundgesamtheit zu als zum Beispiel eine systematische Auswahl (jedes x-te Dossier). So lässt sich beispielsweise in einer Aktenserie des Migrationsamtes ziemlich genau hochrechnen, wie viele Dossiers zu Personen aus der Türkei in der gesamten Ablage vorhanden gewesen sein müssen, wenn man weiß, wie viele entsprechende Dossiers unter B abgelegt sind und wie sich die türkischen Nachnamen über das Alphabet verteilen.

FAZIT

Das Staatsarchiv Zürich hat sich aufgrund dieser Überlegungen dazu entschlossen, in jenen Fällen, wo eine Buchstabenauswahl überhaupt in Frage kommen kann, beim Buchstaben B zu bleiben. Da nicht unbedingt bei allen Aktengruppen eine zehnpromtente Stichprobe angezeigt ist, kann für kleinere Zielgrößen auch nur ein Teil des Buchstabens B übernommen werden. Alle

untersuchten Nationalitäten sind, wenn man ihren Anteil an der Gesamtbevölkerung mitberücksichtigt, in einer B-Auswahl in einem relativ ausgeglichenen Verhältnis vertreten. Obwohl mit einer Kombination von zwei Buchstaben teilweise noch bessere Resultate erzielt werden könnten, ist keine der durchgerechneten Varianten der „B-Auswahl“ so eindeutig überlegen, dass sich eine Umstellung aufdrängen würde. Hier bestätigt sich im Übrigen Matthias Buchholz' Aussage, dass es weder „richtige“ noch „falsche“ Buchstaben gibt.

Dazu kommt, dass es sinnvoll ist, über längere Zeit die gleiche Buchstabenauswahl zu treffen: einerseits um bei den gleichen Aktengruppen längere, vergleichbare Serien zu bilden, andererseits um aktengruppenübergreifende Recherchen zu ermöglichen. Alle Archive, die sich neu für einen Buchstaben entscheiden wollen oder müssen, haben jedoch verschiedene Varianten zur Auswahl.

Thomas Neukom, Zürich

- 3 Matthias Buchholz: Stichprobenverfahren bei massenhaft gleichförmigen Einzelfallakten. Eine Fallstudie am Beispiel von Sozialhilfefällen. In: *Historical socialresearch*, vol. 27 (2002), No. 2/3, S. 100–223.
- 4 Arbeitsgruppe Bewertung des VSA (Verein Schweizerischer Archivarinnen und Archivare): Stichprobenziehung/Sampling. Begriffe und Verfahren (2010). Im Internet unter: http://www.vsa-aas.org/fileadmin/user_upload/texte/ag_koko/Sampling_Begriffe_und_Verfahren.pdf (aufgerufen am 5.11.2013).
- 5 Der Autor bedankt sich insbesondere bei den Mitarbeiterinnen und Mitarbeitern der Abteilung Überlieferungsbildung des Staatsarchivs Zürich für die engagierte Mitarbeit an dieser Untersuchung.
- 6 Buchholz (Anm. 2), S. 130–132.
- 7 Dies entspricht der Berechnung, die auch der Grafik bei Buchholz (Anm. 2), S. 132, zugrundeliegt.
- 8 Buchholz (Anm. 2), S. 131.