The German Surname Atlas Project – Computer-Based Surname Geography

Kathrin Dräger
Mirjam Schmuck
Germany

Abstract

The German Surname Atlas (Deutscher Familiennamenatlas, DFA) project is presented below. The surname maps are based on German fixed network telephone lines (in 2005) with German postal districts as graticules. In our project, we use this data to explore the areal variation in lexical (e.g., Schröder/Schneider ‘tailor’) as well as phonological (e.g., Hauser/Häuser/Heuser) and morphological (e.g., patronyms such as Petersen/Peters/Peter) aspects of German surnames. German surnames emerged quite early on and preserve linguistic material which is up to 900 years old. This enables us to draw conclusions from today’s areal distribution, e.g., on medieval dialect variation, writing traditions and cultural life. Containing not only German surnames but also foreign names, our huge database opens up possibilities for new areas of research, such as surnames and migration. Due to the close contact with Slavonic languages (original Slavonic population in the east, former eastern territories, migration), original Slavonic surnames make up the largest part of the foreign names (e.g., –ski 16,386 types/293,474 tokens). Various adaptations from Slavonic to German and vice versa occurred. These included graphical (e.g., Dobschinski < Dobrzynski) as well as morphological adaptations (hybrid forms: e.g., Fuhrmanski) and folk-etymological reinterpretations (e.g., Rehsack < Czech Režak).

***

1. The German surname system

In the German speech area, people generally started to use an addition to their given names from the eleventh to the sixteenth century, some even later. These additions, the so-called bynames, were at first not inheritable and not fixed, meaning that one person could use different bynames. The term family names can only be used once the names become heritable and fixed. Since the seventeenth century official ordinances were passed to stop changes in family names, to prescribe an obligatory family name and to fix the spelling. Finally, the Civil Code enforced in 1900 made the spelling of family names in Germany permanent. The term surname comprises both the notions of byname and hereditary family name. In the following the term surname is used because it is the most common term (cf. Barker; Spoerlein; Vetter; Viereck 2007: 1).

One can distinguish five groups of surnames according to labelling motivation:

1. Surnames derived from first names, so-called patronymics: e.g., Werner, Walter or Peters. This means that the eponymous ancestor was named after the given name of his father in most cases, but occasionally also after that of his mother, a relative or another person of reference.

2. Surnames derived from professions: The eponymous ancestor was named after his profession, his office or his class. The most frequent German family names belong to this group, such as Müller, Meier, Schmidt.

3. Surnames derived from the provenance of recently arrived persons. In these cases, the eponymous ancestor was named after his land, place or tribe of origin, e.g., Hamburger, Unger, Beier.
4. Similar and sometimes hard to separate are the surnames derived from place of residence: The eponymous ancestor was named after the place where he lived, the surroundings, the site, the material of the house and so on, for example Bach, Berger, Steinhauser.

5. Surnames derived from nicknames: The eponymous ancestor was named after physical, emotional or mental features, after things, plants, animals, and so on, for example Klein, Schwarz, Kluge.

All German family names can be categorized into these groups, but there are so-called meaning differences, i.e., cases in which several labelling motivations led to a single surname.

German surnames preserve linguistic material which is up to 900 years old, from Middle High German, Middle Low German and Early New High German. This enables us to draw conclusions on medieval dialectal variation, writing traditions and cultural life using the current areal distribution.

2. Presentation of the DFA

The high degree of areal variation of the German surname system is now being made accessible by the German Surname Atlas project (Deutscher Familiennamenatlas), a cooperation between the universities of Freiburg and Mainz under the direction of Prof. Dr. Konrad Kunze in Freiburg and Prof. Dr. Damaris Nübling in Mainz. The project started in 2005 and will run till 2012.

The most frequent and impressive examples are selected from the roughly one million different surnames in Germany to answer grammatical and lexical questions. The data basis consists of all fixed network telephone lines in the Federal Republic of Germany in the year 2005 as provided by the Deutsche Telekom AG. To come up with the approximate number of people who bear a specific name one multiplies the number of telephone lines by 2.9. In Germany, telephone lines are the only comprehensive and available data basis. They are arranged by postal code districts consisting of five digits each.

We usually do not create maps with a single name on it but rather maps which combine several names with the same characteristics by using so-called regular expressions. For example, we can illustrate the spreading of the diminutive suffix -le (Map 1) by combining all surnames which end with -le, for example Hämmerle, which means ‘little hammer’. These are roughly 8,700 different names (types) with about 300,000 telephone lines (tokens) altogether. To name another example, you can show the emergence and the spelling of the umlaut with the example Hauser/Häuser/Heuser (Map 2). In this case it does not matter whether the name is written with s, ss or ß, so all of these varieties are combined in groups. As shown on the map, Hauser contains the types Hauser with 6818 tokens, Hausser 114, and Haußer 201. Häuser contains Häus(s)er 1362+185, Häußer 438, Haeus(s)er 21+7, and Haeußer 3. Heuser contains Heus(s)er 3395+50, and Heußer 110.
Map 1. All German names with diminutive suffix -le combined.
Map 2. Emergence and spelling of the umlaut in Hauser/Häuser/Heuser.

It is possible to create relative maps, which show the percentage of all telephone lines in the corresponding postal code district, or absolute maps, which show the actual number of telephone lines in the postal code district. Absolute maps are only convenient for surnames with less than about 1,000 tokens because the major urban centres produce the biggest spots, as the example of the surname Richter (59,950 tokens) shows (Maps 3 and 4). One can use circular charts, like those on Maps 3 and 4, or areal charts (Map 5). One can also combine both as shown on Map 6, which shows the spreading of the apocope of the e in the nickname Lange/Lang. Indicated is not only the spreading with the circular charts or the colouring of an area, but also the density with the size of the circular charts. The maps are adjustable to one-, two-, three- or five-digit postal code districts. In most cases it has proved most suitable to use three-digit postal code districts like on Map 7, which shows the spelling of the profession name Schmitz/Schmitt/Schmi(e)d. Five-digit maps are used to show the spreading of less frequent names or to show details (Map 8).

The atlas will contain two parts, one with grammatical topics and one with lexical topics. The first part, containing phonology, graphemics, morphology and syntax, will be published in two volumes; the first volume will be ready in 2009. The second part will be divided into different sections based on the five surname types: patronyms, profession names, provenance names, residence names, and nicknames.

Each map is accompanied by a detailed commentary containing six sections: We start by describing which topic is being illustrated and why this special case has been chosen. Usually,
very frequent names are selected which are preferably etymologically unambiguous and deliver a convincing map. Secondly, the quantitative database for the map is indicated with the regular expression applied, the output types and the frequencies of the different types. The third section contains etymological information about the names. Section four comprises further details about the map and the so-called Nebenkarten. These smaller maps contain details from the main map or illustrate the same topic with other examples. In section five, historical forms of the names are listed. The German Surname Atlas is the first linguistic atlas which considers data both from the present and from the past reaching back to the Middle Ages. The last section contains bibliographical references, cross references and further information, for example the frequency and spreading of the names in neighbouring countries.

Map 5. Areal chart (relative) for Richter.
Map 6. Apocope in *Lange/Lang*. 
Map 7. Distribution of Schmitz, Schmitt, and Schmi(e)d, three-digit postal code districts.

Map 8. Distribution of Schmitz, Schmitt, and Schmi(e)d, five-digit postal code districts.
3. Case study: *Hermanski, Szulczynski* etc. – Germano-Slavic hybrids

Our huge database contains not only German surnames but also foreign names and in particular hybrid names. As a result of the close contact with Slavonic languages, original Slavonic surnames and Germano-Slavic hybrids make up the biggest part. Hybrid names, such as *Hermanski, Szulczynski* etc. will not be taken into account in the DFA. Nevertheless, they constitute a surprisingly large number of German surnames. Due to computer-based research they are now available in great variety and thus provide a new interesting field of research.

3.1 Slavonic names in Germany

According to Naumann (2007), 15% of surnames in his data are Slavonic in origin. This fact is a result of the long language contact between German and West Slavonic languages, especially Polish and Czech. Intensive language contact is above all due to the former eastern territories of the German Reich and the Slavonic settlement in the eastern part of Germany. Further Slavonic influence is due to the displacement of the German partly assimilated population in the eastern territories after the Second World War and their return to Germany. Last but not least, there have been a great number of Polish immigrants especially in industrial areas, above all in the Ruhr area. Those migrant labourers came mainly in the middle of the 19th century (cf. Hengst 2001; Jachnow 1991/92; Rymut/Hoffmann 2006: VII–XIV; Schmalz-Jacobsen/Hansen 1995: 385–401, 501–511).

Map 9. Polish names in -ski 16,386 types/293,474 tokens
To illustrate the great number of Slavonic names, Map 9 shows the distribution of all surnames ($\geq 2$ tokens) ending in the typical Polish suffix -ski (16,386 different names, total 293,474 tokens). As we would have expected from the historical background Polish names are concentrated above all in the Ruhr area, with a lower concentration of names ending in -ski distributed in the northeast. Note their high token-frequency. Map 10 shows the distribution of names with the Czech variant of the suffix (4,408 types/41,535 tokens). Those names are concentrated in the Sorbian area and in southeastern Germany.\\footnote{1}

Because of the historical background, we have to deal with two kinds of language contact: German names in Slavonic speaking surroundings and Slavonic names in Germany. Likewise, distinction has to be made between the two directions of adaptation: from Slavonic to German and vice versa.\\footnote{2}

### 3.2 Levels of adaptation

Adaptations occur on different levels: four levels can be distinguished (adaptations of Polish names in the Ruhr area cf. Burghardt 1975; Menge 2000):

1. Graphical
   
   \[
   \text{Szymanski} \rightarrow \text{Schimanski}, \quad \text{Kubic} \rightarrow \text{Kubit},
   \]
   
   \[
   \text{Schulz} \rightarrow \text{Szulz}, \quad \text{Friedrich} \rightarrow \text{Frydrych}
   \]

2. Phonological
   
   \[
   \text{Brzeski} \rightarrow \text{Breski}, \quad \text{Majchrzak} \rightarrow \text{Maischak}
   \]
3. Morphological

\begin{itemize}
\item Blazek > Blaschke, Kubik > Kubig
\item Baier > Bajerski, Schulz > Szulinski
\end{itemize}

4. Lexical, i.e., translation or folk-etymological reinterpretation:

\begin{itemize}
\item Dombrowski > Dombrück, Majchrzak > Maischatz
\end{itemize}

Below we will take a closer look at morphological adaptations, especially cases such as Bajerski, where German surnames are adapted to Slavonic and second folk-etymological reinterpretations of original Slavonic names, the case Majchrzak > Maischatz.

3.3 Morphological adaptations

First of all some short remarks on the Polish surname system: Polish surnames are characterized by their typical suffixes, the most common is -ski, reflected to an equal degree in our data.

Originally the suffix -ski is added to property names (Tarnów > Tarnowski ‘owner of Tarnów’). Thus -ski was typical for surnames of the nobility. Later on the suffix occurred in provenance names (Kraków > Krakowski). Due to its enormous prestige, -ski came off as an onymical suffix and was no longer restricted to provenance names, but combined with every base (Kowalski < kowal ‘smith’, Szymanski < Szymon, Simon). The suffix -ski is still extremely frequent: Among the 100 most frequent Polish surnames 65 end in -ski (cf. Skowronek 2001: 186f.). Often the suffix is combined with other suffixes: -owski (Kwiatkowski), -ewski (Wisniewski), -inski (Dobilinski).

Other frequently used Polish suffixes:

- *-ak* suffixes (= diminutive suffix): -ak (Nowak), -ek (Kowalek), -ik (Wojcik), -czak (Ideczak), -czyk (Jaworczyk)

- *-icz* suffixes: -icz (Panicz), -owicz (Jurakowicz), -ewicz (Chudakiewicz)

Those Polish suffixes have also been used to “slavonize” German surnames. We have chosen the example Fuhrmann which means ‘waggoner’ (Table 1). In our data there are 41 different names altogether with a total of 547 tokens.

The most frequent suffix is -ek (12 types/269 tokens), followed by -ski (14 types/161 tokens). Other, less frequent suffixes, are -ak, -czyk and -wicz. All these suffixes are Polish: the suffix -ek could be either Polish or Czech.
Table 1. Hybrid forms of the German surname Fuhrmann ‘wagoner’

Table 2 (see appendix) shows the large number of hybrid German surnames with Polish *-ski* suffixes. Taken into account are all names with a minimum of 5 tokens. In our data there are 38 different names altogether (870 tokens) with simple *-ski* suffix. Combinations occur more often. The most frequent suffixes are *-owski* (45 types/922 tokens) and *-kowski* (31 types/396 tokens). Other combinations are *-ewski* or *-czewski* and *-inski* or *-czynski*. Similar to Polish surnames, *-ski* suffixes occur with all five name types: profession names (*Lehmanski* < ‘tenant’, *Schuttkowski* < ‘marksman’), patronyms (*Hermanski*, *Willamowski*), provenance or residence names (*Bajerski* < *Baier* ‘person from Bavaria’, *Straßewski* < *Straßner* ‘person who lives near the street’) and nicknames (*Grotkowski* < *Grot(e)* ‘tall person’, *Schwartinski* < *Schwarz* ‘black haired person’).

### 3.4 Folk-etymological reinterpretations of Slavonic surnames

A common way of adapting Slavonic names to the German system is folk-etymological reinterpretation. We have to deal with two kinds of adaptation: semi- and total-adaptation. Semi-adapted forms are cases such as *Korczak* > *Korschatz* (*Kor* ‘?’ + *Schatz* ‘treasure’) and *Kierzkowski* > *Kirschkowsk* (*Kirsch* ‘cherry’ + *-owski* ‘?’). Examples for totally adapted forms are *Kierzkowski* > *Kirschkopf* (*Kirsch* ‘cherry’ + *Kopf* ‘head’) and *Melczak* > *Mehlsack* (*Mehl* ‘flour’ + *Sack* ‘sack’). Examples in tables 3 and 4 show the folk-etymological reinterpretation of the frequent Slavonic suffix *-czak*. Table 3 illustrates the adaption to the German word *Schatz* ‘treasure’. The first part of the table concerns cases of total adaption to German (*Mannschatz*, *Bauschatz*, *Maischatz*, *Goldschatz* etc.), where both elements of the compound are fully transparent. In the second part of the table semi-adapted names are subsumed such as *Buschatz*, *Waschatz* and *Marschatz* where the first element is opaque.
<table>
<thead>
<tr>
<th>Profession names</th>
<th>Patronym: (&lt; Germanic first names)</th>
<th>Provenience/residence names</th>
<th>Nicknames</th>
</tr>
</thead>
</table>

Table 2. German surnames with Polish -ski suffixes (< 5 tokens)
### Table 3. Folk-etymological reinterpretation of slav. -czak > -schatz ‘treasure’

Table 4 exemplifies the adaption of Slav. -czak to Germ. Sack ‘sack’. The DFA-data contains compounds such as Maisack, Brotsack and Milchsack, Rehsack. All of them are totally transparent, but only a few of them make sense, for example Mehlsack ‘flour sack’. Obviously most of these names represent nonsense compounds (Rehsack, Maisack). This name type calls to mind bipartite Swedish names such as Stenqvist ‘stone twig’, Malmqvist ‘ore twig’, etc. Those names are in some way ideal proper names: They are distinct from common nouns without being opaque (cf. Nübling 2000).

<table>
<thead>
<tr>
<th>Adapted names ending in -schatz (database DFA)</th>
<th>Possible Slavonic base</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man(n)schatz ‘man’-</td>
<td>15+59</td>
</tr>
<tr>
<td>Bauschatz ‘building’ + ‘treasure’</td>
<td>40</td>
</tr>
<tr>
<td>Buchschatz ‘book’-</td>
<td>17</td>
</tr>
<tr>
<td>Maischatz, Meischatz ‘May’-</td>
<td>25+10</td>
</tr>
<tr>
<td>Goldschatz ‘gold’-</td>
<td>2</td>
</tr>
<tr>
<td>Bockschatz ‘buck’-</td>
<td>9</td>
</tr>
<tr>
<td>Buschatz</td>
<td>45</td>
</tr>
<tr>
<td>Waschatz</td>
<td>24</td>
</tr>
<tr>
<td>Marschatz</td>
<td>12</td>
</tr>
<tr>
<td>Korschatz</td>
<td>3</td>
</tr>
<tr>
<td>Laschatz</td>
<td>2</td>
</tr>
<tr>
<td>Wierschatz</td>
<td>2</td>
</tr>
</tbody>
</table>

- Manczak (< Manomir)
- Buczak (< buczeć ‘cry, grumble’)
- Buchczak (< buchta ‘piece of bread, cake’)
- Majczak, Maichrzak (< Melchior)
- ?Golczak (< goly ‘bare, fundless’), ?Kolczak (< Nicholas)
- ?Boguslav
- Grze(sz)czak (< Grzegorz, Gregor)
- Buczak (< buczeć ‘cry, grumble’)
- Waczak (< Waclaw)
- Marczak (< Marek)
- Korczak (< korczak ‘wooden cup’)
- Laszak (< Lazarz, Lazkarz)
- Wierzak (< wiara ‘faith’, Veronika)
Table 4. Folk-etymological reinterpretation of Slav. -(c)zak > -sack ‘Sack’

4. Conclusion

Computer-based surname geography enables us to systematically register all surnames in their enormous variety. Thus even names with fewer tokens (e.g., hybrid names) can be taken into account and etymologies can be verified. In addition to anthroponomastics, surnames provide a huge database for further interdisciplinary research. Containing linguistic material from Middle High German, Middle Low German, and Early New High German, German surnames represent an important database for historical linguistics as well. Furthermore, surnames provide a rich source for other fields of research, e.g., migration and all kinds of historical, social or cultural studies, genetics and many more.

Notes


2. Farkas (this volume) provides a similar approach for the situation in Hungary. He points out that influence is exerted both by Hungarian surnames on foreign names (e.g., German or Slavonic names) and by foreign names on the dominant Hungarian system.

References


Farkas, Tamás (this volume). Surnames of foreign origin in language contact situation: the reasons and ways of their changes and their influence on the surname stock in Hungary.


Mirjam Schmuck
Johannes-Gutenberg Universität
FB 05 Deutsches Institut
55099 Mainz
GERMANY
mschmu@uni-mainz.de

Kathrin Dräger
Albert-Ludwigs-Universität
Deutsches Seminar I
79085 Freiburg
GERMANY
kathrin.draeger@germanistik.uni-freiburg.de