

**From Brzezinski to Brzezinski:
How European Immigrants Kept Their Last Names**

Josh Goldstein
Department of Demography
UC Berkeley
josh.goldstein@berkeley.edu

Guy Stecklov
Department of Sociology
University of British Columbia
guy.stecklov@ubc.ca

Short abstract:

Family name changes are a staple of immigration assimilation lore, but such reports are based on those who did change their names and not on the total population of immigrants. In this paper, we use a new source to look for the first time at population-level records that show both those who retain their names and those who changed them. Social Security applications from the 1930s to the 1960s allow comparison of the last names of immigrant men and their fathers. We find the vast majority of immigrant sons kept their family name, even when last names were difficult for English speakers. We use language classification methods to categorize changes, and find that most changes preserved ethnic recognizability. In this light, the so-called "re-emergence" of ethnicity later in the century was less uncovering something hidden than a renewed emphasis of ethnic origins kept in plain sight.

Extended abstract:

The point about the melting pot, as we say later, is that it did not happen.

-- Preface to Beyond the Melting Pot (1963,
page v.) by Glazer and Moynihan

Immigrants arriving from Europe in the early part of the 20th century faced a wide array of pressures to become "American." From learning English, to serving in the military, to becoming good citizens, immigrant allegiances and identities were not supposed to be mixed but rather, in the words of Teddy Roosevelt "American and nothing else."

It is often remarked that anglicization of last names was a revealing and important part of Americanization. This transformation of a surname's original form is reported as an important part of many Americans "origin story." There is also a rich literature based on archival research of name-changing petitions, revealing the motivations of those who

wanted to change their names. This literature emphasizes the discrimination that immigrants faced and the pressure to assimilate into the "melting pot" as well as the change in attitudes that resulted in the "re-emergence" of ethnicity in the 1960s and 1970s.

However, a serious problem with viewing name changes through the lens of those who changed their name is that it is blind to those immigrants, and their descendents, who chose to retain their name. In this study, we take advantage of a newly available source, full-population administrative records from Social Security from the 1930s through 1960s, in which applicants reported both their own legal name and that of their parents. We focus primarily on the names of immigrant men and their reports of their father's names - name changes among women are more complicated to study because of norms that vary across ethnic groups for adopting spousal surnames. Our approach allows us to see both those who kept their names and those who changed them. We also use language classification models to distinguish between different kinds of name changes, specifically those that are "Americanizing" (e.g., Brzezinski to Brown) and those that are "neutral", retaining the recognizability of an ethnic origin (e.g., Brzezinski to Brezinski).

Our population-level study reveals (1) that name changes tended to be the exception rather than the rule, with more than 90% of immigrant men reporting that their name was identical to their fathers. Furthermore, (2) we find that of those who did change their names, the majority of these changes did not hide the ethnic origin of the applicant. The share of immigrants that actually changed their name to something more "English" was small. (As we report in preliminary results below, our current estimates are that only about 1% of Italians, about 2% of Greeks, 3% of Germans and Russians, and 5 to 6% of Hungarians and Austrians changed their names from recognizably ethnic names to recognizably English names).

Our study has several potential implications for how we think about the American immigrant experience in the early and mid 20th centuries and the re-emphasis on ethnic identity that emerged later as well as on the methodology of using administrative records for population analysis and use of modern natural language classification methods.

- Although there was clearly pressure to assimilate, this did not encourage most people to abandon their recognizably ethnic names. Most immigrants left their names completely unchanged, and most of those who did change their names did not erase the recognizability of where they came from. As such the reality of the early and mid- 20th Century was not so much a "melting pot" in which origins

were hidden, but rather one in which immigrants of recognizably ethnic origin became American in other ways.

- The so-called "re-emergence of ethnicity" in the 1960s and 1970s was less the uncovering of hidden origins than a renewed emphasis on origins that were kept in plain sight. Arguably, immigrants never lost their identity of origin; rather at different periods it was thought to be less and then perhaps more important.
- It is important to use population-level analysis as a complement to focused archival research in order to get a representative impression of the immigrant experience. The archival research can reveal the motivations, in this case, of those who changed their names but is not informative about the frequency of name changes or the motivations of those who did not change their names.
- Methodologically, advances in language classification algorithms now makes it possible to categorize name changes of different kinds. Here we distinguish between "Americanization" and "neutral" kinds of name changes. Future work could attempt to uncover more nuanced variation in the types of name changes occurring for these populations.

Data

We rely on publicly available Social Security application records of those who died from 1988 to 2005 that were released by the National Archives. These records are available as part of Berkeley Unified Numident Database (BUNMD). The individual records include the applicants legal first and last names, along with those of their father and mother. We focus our attention on foreign-born applicants ("immigrants") from European origins with a substantial number of immigrants applying for Social Security cards from 1936 to 1960. These countries along with the number of applicants in BUNMD are given in the table below.

Birthplace Count

Austria	4593
Denmark	2323
England	24864
Finland	1042
France	4177
Germany	30914
Greece	7828

Hungary	9310
Ireland	11335
Italy	39728
Netherlands	5694
Norway	4634
Poland	28674
Russia	15368
Sweden	2782
Total	193266

Table 1: Numbers of immigrants in the Berkeley Unified Numident Data Base who applied for social security cards from 1936 through 1960 (Note: Our sample excludes a small number of individuals who did not report father's names.)

The principal advantage of this source for studying name changes is that it records both child and parent's names on the same record, rather than relying on linking records nominally, which itself can fail when names change.

A disadvantage of the common reporting by a sole applicant of father and sons names is that there may have (1) been a tendency to either write the same name or transcribe the same name, even if the names used by father and son in their actual lives differed and (2) some immigrant fathers names appear to have themselves been altered, so that for example a son born in Poland reports both his own and his father's name as "Smith" (the translation of the common Polish name "Kowalski").

Another disadvantage is that Social Security applications do not give age-at-entry into the United States for immigrants. This can be addressed partially by looking at age-at-application for Social Security. Those who apply at very young ages (e.g. 18) tend to be those who immigrated as children and waited until age 18 to apply, whereas those who applied at older ages (e.g., 35) tended to be more recent immigrants.

Methods

Cleaning:

We process these data using a series of steps that clean names. This includes removing non-alphabetic characters and removing suffixes like JR and SR.

Classification:

The first stage of our classification distinguishes between "changers" and "non-changers" based on the exact match (or lack thereof) in the reported character strings of father and son's names.

In the second stage, we distinguish between two kinds of name changes: Americanizing changes that hide the ethnic origin of the name (e.g., Lippschitz --> Lipton or Schmidt --> Smith) and "neutral" changes that retain the recognizable ethnic origin of an immigrant (e.g, Golsteyn → Goldstein or Papadopoulus --> Pappas).

There are many possible methods for classifying between Americanizing and neutral changes, including using human coders, supervised learning approaches based on features of the name (name length, vowel endings, etc., the use of dictionaries to see if names appear more frequently say among English immigrants than among Italian immigrants. For now, we are using a well-known language classifier "textcat" that uses n-grams (the frequency of strings of varying length, 1, 2, 3, 4, and 5 characters) to assign languages (Hornik et al. 2013).

To implement the classifier, we produce a language profile for each immigrant country based on the names of several thousand reports of father's names in the SS-5 records. This produces a "profile" n-gram distribution. We also do this for all immigrants from England. We then classify each name of the son in terms of the n-gram distribution of that name and whether it more closely resembles the profile of the country of origin or that of England.

The advantage of this n-gram method is that it is possible to apply to every name, even if it only appears once for the sons and not at all for the fathers. A disadvantage is that the English sample may not contain exclusively "English" origin names as some of those who immigrated to the United States from England had names from other origins, for example Italy or Poland or Russia. Another challenge is that immigrants sometimes have names that are not typical of their country of origin or of England (for example, an Austrian immigrant carrying an Italian name) and in this case the binary classification between England vs. Austria is not easy.

No change:	Papadopoulus --> Papadopoulus
Neutral change:	Papadopoulus --> Pappas
Americanizing change:	Papadopoulus --> Pope

Table 2: Examples of name change classification

Preliminary Results

As a first cut on our data we looked at selected names, choosing common names for each country of origin, and looking only at whether there was any change in name between father and immigrant son. Figure 1 shows the high rate of retention of last names by immigrants for selected ethnically distinct last names. We can see that there is a low level of "background changes," even for already English names like "Smith" and "Johnson", with a few percent of sons reporting different names. Other highly distinctive ethnic names like Mueller or Rossi however do not appear to have significantly larger rates of change. Jewish names like Cohen, Rosenbaum, etc.. do appear to have somewhat lower retention rates; and surprisingly Scandinavian names like Johansson have the lowest retention rate, with a very high rate of changes. Many of these are purely orthographic (from Nillson to Nilson) but some (Nillson to Nelson) could be considered Americanizations.

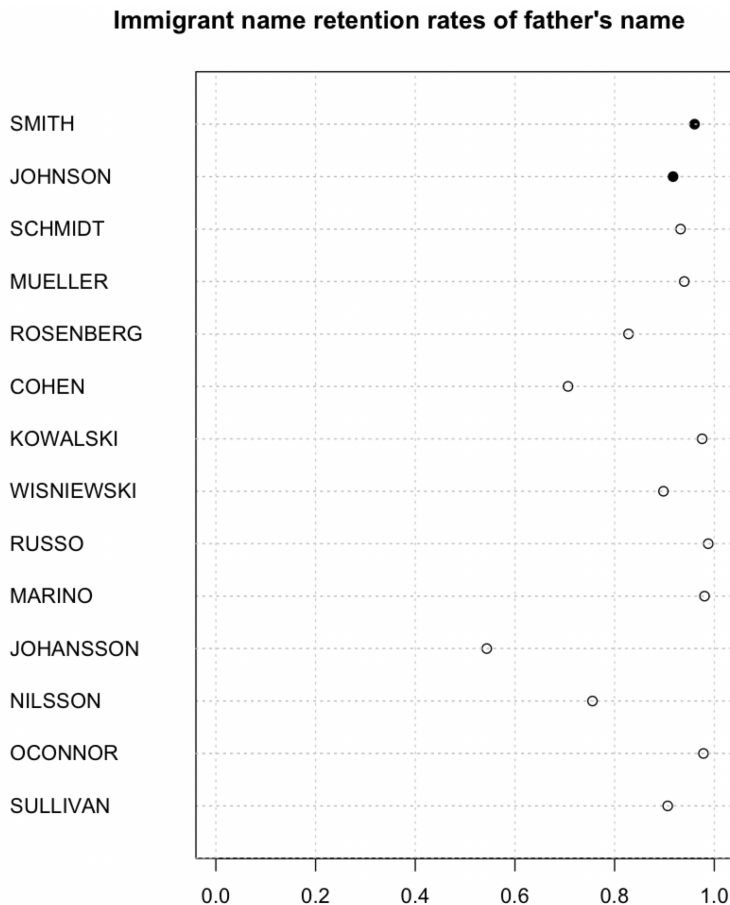


Figure 1: Name retention rates by selected popular last names typical of several origins. Smith and Johnson are shown for comparison. Immigrants can be from any of the foreign countries in the study.

The second figure shows change rates by the country of birth of immigrants and applies our language classification algorithm to distinguish between neutral and americanizing changes. We do not apply this classification algorithm to Irish and English names since we do not have any non-English profile to compare with. For all other countries, we compare the n-gram distance to the English profile with that of the profile of the origin country. We find that for most countries only a very small share of those name changes actually result in a name that is classified as "English" rather than the country of origin. For Italy, Netherlands, France, Norway and (surprisingly) Greece, the classified portion of Americanizing name changes was 2 percent or less. For Germany, Denmark, Russia, Finland, Sweden, and Hungary, the fraction was 3 to 4 percent, and for Poland and Austria the changes were 5 and 6 percent.

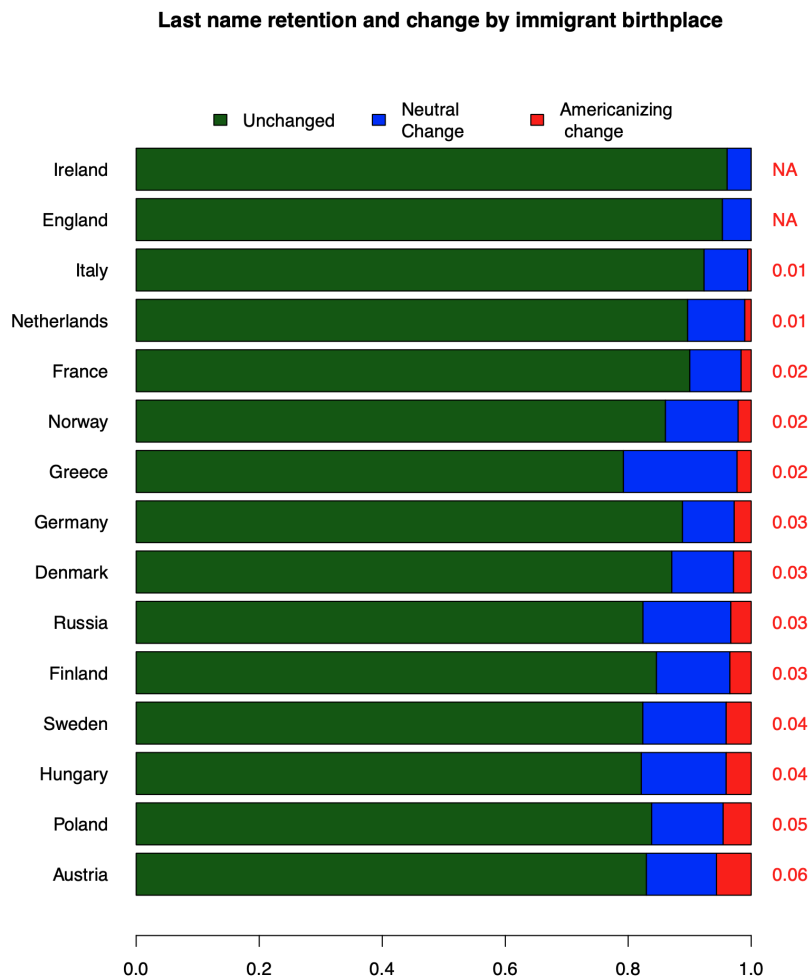


Figure 2: Last name changes and retention by immigrant birthplace, as coded by textcat language classifier based on n-gram profiles of immigrant names from each profile, compared to the profile of English immigrants. Ireland and England are not classified. Red numbers show Americanization rates.

Future Work

For the complete paper (which we hope to present at EPC) , we plan to

- Improve our analysis by
 - a. Experimenting with alternative classification methods, including asking humans to classify name changes and then using supervised learning techniques like random forests to imitate human classification. We have explored supervised machine learning on a small sample and this method appears to have promise.
 - b. Quantify the extent to which the SS-5 records miss early name changes that are applied to father's last names. One approach for doing this is to measure the share of immigrant father's last names that are classified as "English".
- Expand our analysis to include
 - a. Name changes among the native born who report ethnic names for their fathers. This will allow us to see whether there were higher or lower name change rates for later generations, many of whom had assimilated in other ways.
 - b. Comparison of European immigrant name change rates with immigrants from non-European origins such as Japan, China, Mexico, the Philippines, Jamaica and Haiti.
- Explore variation in name changes
 - Over time
 - By age, and years since arrival
 - By SES (including education and income)
 - By context (state of residence and perhaps smaller geographic units)
 - For different ethnic origins from a single birth country (e.g., Jews and non-Jews)

More ambitiously, we may use our language classifier to look at all immigrants in the full-count 1930 census to see who and where immigrants appear to be Americanizing their names. This source also has age-at-entry.

Selected Bibliography

Breen C and Goldstein JR. Berkeley Unified Numident Mortality Database: Public administrative records for individual-level mortality research, *Demographic Research* 47(5):111–142 (2022)

Hornik, K. et al. The textcat Package for n-Gram Based Text Categorization in R. J. Stat. Softw. 52, (2013).

Fermaglich, K. *A Rosenberg by Any Other Name*. (NYU Press, 2018).

Glazer & Moynihan, D. P. *Beyond the Melting Pot: The Negroes, Puerto Ricans, Jews, Italians, and Irish of New York City*. vol. 63 (MIT Press, 1963).