

## Long-term sociolinguistics trends and phonological patterns of American names

Katherine He\*

**Abstract.** This paper identified macro trends and phonological patterns of 348 million American baby names over 137 years from 1880 to 2017. The analysis showed that sociolinguistic trends have significantly influenced naming over time, as seen in the rise of individualism and unisex names, the impact of public figures and pop culture, and the substantially higher count of unique female names compared to male names. In addition, phonological analysis showed significant differences between male and female names in the number, type, and location of vowels as well as the number of syllables. On average, female names had more vowels, less consonants, and more syllables than male names. Also, names with certain word-final vowels and consonants were identified to be mostly-female or mostly-male. These findings demonstrated an inherent correlation between phonology and the perceived gender of names.

**Keywords.** onomastics; American names; long-term sociolinguistics trends; phonological patterns

**1. Introduction.** Used to identify individuals, names are associated with traits that influence how people perceive one another, which in turn affects naming trends. This research examined this phenomenon on a large scale, as demonstrated by trends and phonological patterns in American baby names over 137 years from 1880 to 2017.

Social perception is becoming increasingly important in today's interconnected world. A person's name constitutes a large part of personal and social perception of identity, as having a name is the first step toward differentiating one person from another.

Previous studies have shown how names could influence personal identity. The correlation between names and perceived identity, name stereotypes, was explained in a 2017 study (Zwebner et. al, 2017) that explored how appearance could be shaped by one's name. In addition, previous research has suggested that names and their phonology are correlated with personal development and life decisions due to implicit egotism (Pelham et. al, 2002).

Therefore, given that names have an influence on identity, it is not surprising that names influence social perception as well. For example, one study (Nick, 2017) identified covert name biases by examining how the same essays with different names would score differently. Another study (Bertrand & Mullainathan, 2004) showed that how some minority-sounding names on fake resumes had substantially lower callbacks than white-sounding names on the same resumes. Other research has investigated sociolinguistics trends through the lens of names – either how names could be affected by cultural trends (Twenge & Campbell, 2010) or be modelled as an evolution (Hahn & Bentley, 2003).

While past research explored the correlation between names, identity, and social perception, this research focused on the macro trends of naming over a long period of time (137 years, from 1880 to 2017). In addition, this research studied the phonological patterns of American names by

---

\* Appreciation for the support of Professor Stephanie Shih, graduate student Jessica Johnson, and the Language and Computation (ShLAC) Lab at the University of Southern California, as well as Science Research teacher Melissa Klose of the Palos Verdes Peninsula High School. Author: Katherine He, Palos Verdes Peninsula High School (khe@fadingvoices.org).

transcribing the dataset into International Phonetic Alphabet (IPA). While before a connection was made on a smaller scale (Barry & Harper, 1995), this research demonstrated an inherent correlation between phonology of the names and the perceived gender on a macro level.

**2. Data preparation.** The raw data came from the United States Social Security Administration (SSA, 2018). The dataset contained first names for 348 million American babies born over 137 years from 1880 to 2017 and was organized as individual files for each year.

Python programs were written to compile the individual files into a master file and to conduct various analyses. Table 1 shows, compiled in the master file, the total number of babies born (~348 million), the total number of data records (~1.9 million), and the total number of unique first names (~97 thousand) broken down by gender from 1880 to 2017.

| 1880-2017         | Male        | Female      | Total       |
|-------------------|-------------|-------------|-------------|
| # of Babies Born  | 175,749,438 | 172,371,079 | 348,120,517 |
| # of Data Records | 786,372     | 1,138,293   | 1,924,665   |
| # of Unique Names | 40,927      | 67,046      | 97,310      |

Table 1. Total number of babies born, data records, and unique names, by gender, 1880-2017

To study phonological patterns, all names were then transcribed into the International Phonetic Alphabet (IPA) using three IPA dictionaries sequentially: the Amerlex dictionary, the Carnegie Mellon University (CMU) dictionary, and the LOGIOS Artificial Intelligence tool.

A Python program was used to automate the transcription process and to add the IPA transcriptions to the master file. Names were first run through the Amerlex dictionary. The names that were not found in the first dictionary were then run through the CMU dictionary. The remaining names were then run through the LOGIOS tool, which transcribed the names using artificial intelligence. Names transcribed by LOGIOS were hand combed in order to ensure consistent transcription with the other two dictionaries.

The final master file was structured to include the name, gender, popularity (number of babies born), the year of birth, and the IPA for each record.

**3. Analysis findings.** All measurements were normalized against population to remove the effect of population growth. After normalization, analysis shows that the number of unique names per thousand people has grown significantly over time (Figure 1). Although the chart appears to illustrate the trend first decreasing before increasing, this can be attributed to the limitation of data before 1935 (when birth name collection became mandatory for Social Security reasons). Especially during the births of the Baby Boomer generation from 1946-1964, Figure 1 clearly shows that more and more unique names were being introduced into the US population, eventually increasing the number of unique names per thousand people from about 2 to 8 for males (a 4x increase), and from 4 to 11 for female (a 2.75x increase). This suggests that novel, non-conventional names have become more popular over time, which is likely driven by the rise of individualism.

Figure 1 also shows that, except for pre-1935 years when the data was limited, there are significantly more unique names for females than males. For example, currently females have on average about 11 unique names per thousand, which is 37% more than males with only 8 unique names per thousand. This shows that Americans are more open to give females novel, non-conventional names and giving the males more existing, conventional names.

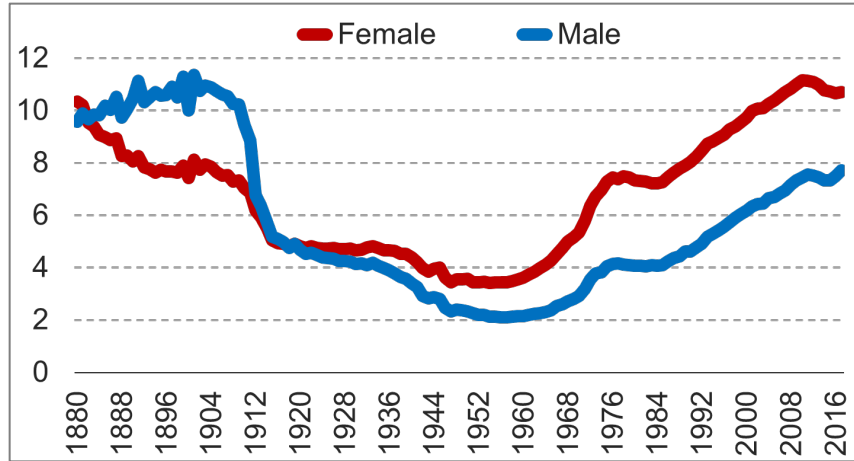


Figure 1. Number of unique names per thousand people, by gender, 1880-2017

The data shows that many names are used for both males and females, which are also known as unisex names. Figure 2 shows that, when normalized against population, there is clearly significant growth of the number of unisex names, especially starting around the 1950s. This demonstrates a growing trend of naming babies without gender association, which can also be interpreted as a form of individualism.

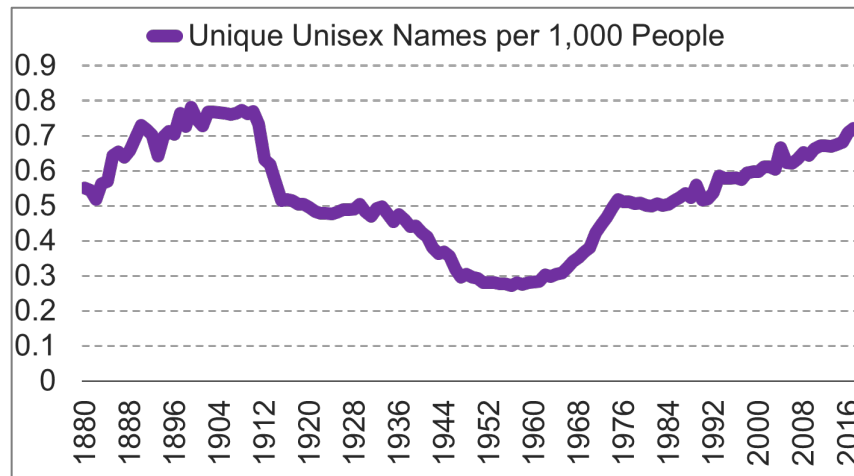


Figure 2. The rise of unisex names per thousand people, 1880-2017

While it is expected that public figures and pop culture can influence social perception of particular names, and thus naming trends due to the mere-exposure effect, this study reviewed two specific datasets with good control.

The first sample was names of US presidents. The analysis looked at the ratio of US presidents' name counts, both 4 years before and 4 years after the presidents' first election. The result (Figure 3) shows that names that are either associated with popular presidents and/or unusual first names (e.g. president Woodrow Wilson and president Barack Obama) had a clear jump (40x and 25x) in the number of corresponding baby names after the presidents' first election.

The second sample was names of Disney movie protagonists. The analysis reviewed the ratio of Disney protagonists' names 5 years before and after the movies premiere (Figure 4). All name counts are normalized to start at 100% at 5 years before movies' premiere. Similar to the president study above, the popular and/or unusual protagonists' names gained significant in-

crease in the newborn population after the movie. For example, the name “Belle” from Beauty and the Beast released in 1991 saw almost 900% increase in name count over the 5 years after the movie premiere.

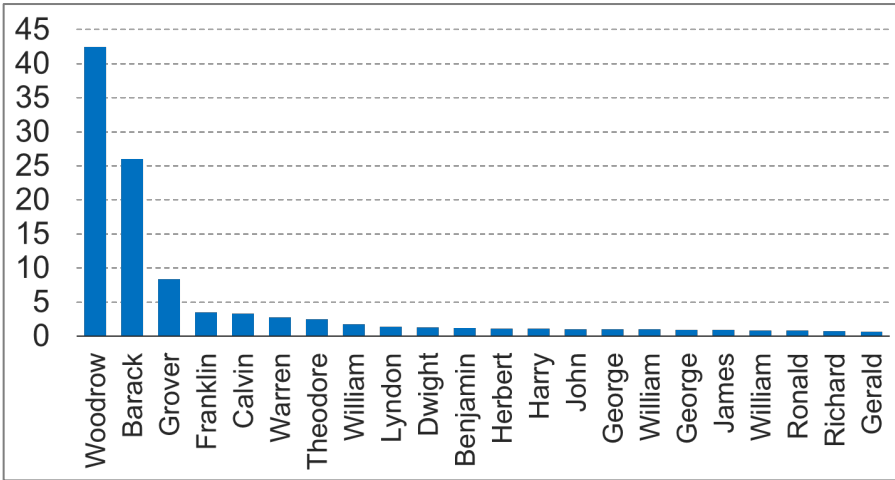


Figure 3. Ratio of US presidents’ name count, 4 years before and 4 years after presidents’ first election

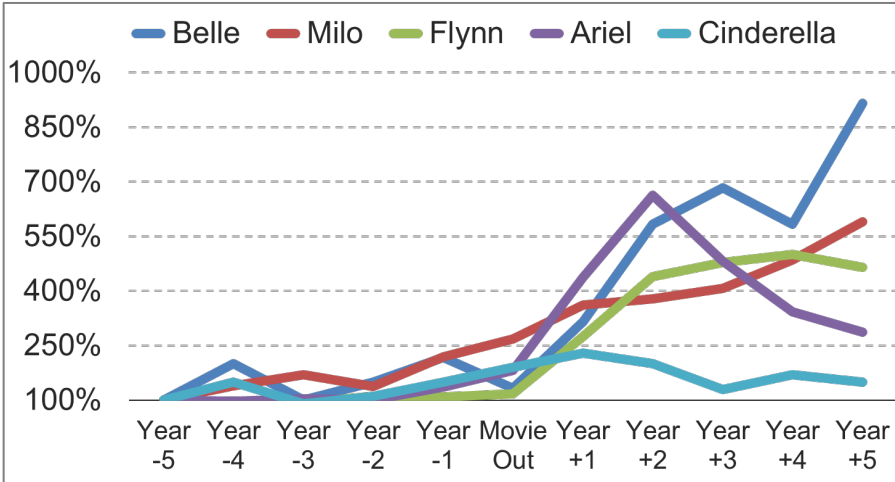


Figure 4. Impact of pop culture: Disney protagonists’ names popularity, 5 years before and after the movies premiere

Phonologically, the analysis shows significant differences between male and female names. Figure 5 shows female names on average have 25% more vowels than male names, while male names have 8% more consonants.

In addition, female names end in vowels for 37% of all females in the dataset, while male names only end in vowels for 8% of the males. Similar difference is also seen in names beginning with vowels, though the percentages are much smaller, with only 8% of female names and 3% of male names starting in vowels (Figure 6).

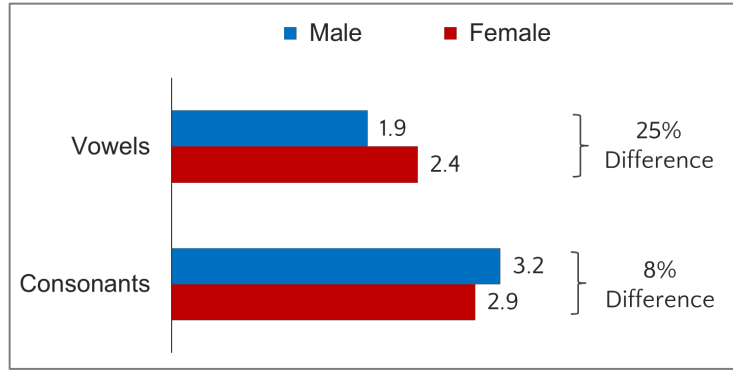


Figure 5. Average number of vowels and consonants per name, by gender, 1880-2017

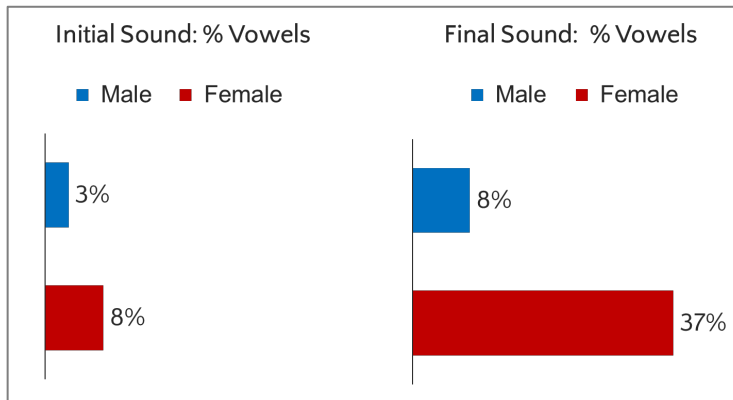


Figure 6. Percentage of names with initial or final sounds as vowels, by gender, 1880-2017

Interestingly, when names end with certain vowels, they are mostly female. Table 3 shows that in the dataset, there were 17 million names ending with [ʌ], of which 97% are female, and there were 49 million names ending with [ə], of which 94% are female.

Conversely, when names end with certain consonants, they are mostly male. Table 3 shows that in the dataset there were 3 million names ending with [f], of which 99% are male, there were 9 million names ending with [k], of which 95% are male, and there were 18 million people's names ending with [d], of which 94% are male.

| Mostly-Male Ending Consonants |                                       |                             | Mostly-Female Ending Vowels              |                                       |                             |
|-------------------------------|---------------------------------------|-----------------------------|--|---------------------------------------|-----------------------------|
| IPA                           | % of Total Population (Male & Female) | Total Population (Millions) | IPA                                      | % of Total Population (Male & Female) | Total Population (Millions) |
| [f]                           | 99%                                   | 3                           | [ʌ]                                      | 97%                                   | 17                          |
| [k]                           | 95%                                   | 9                           | [ə]                                      | 94%                                   | 49                          |
| [d]                           | 94%                                   | 18                          | <i>(n/a: no other dominating vowels)</i> |                                       |                             |

Table 3. Mostly-male ending consonants and mostly-female ending vowels, 1880-2017

Male and female names also differ in syllable count. As shown in Figure 7, of all female names from 1880 to 2017, 29% have three-syllable names compared to 11% for male names. On the other hand, 21% of male names have only one syllable, compared to 6% of female names.

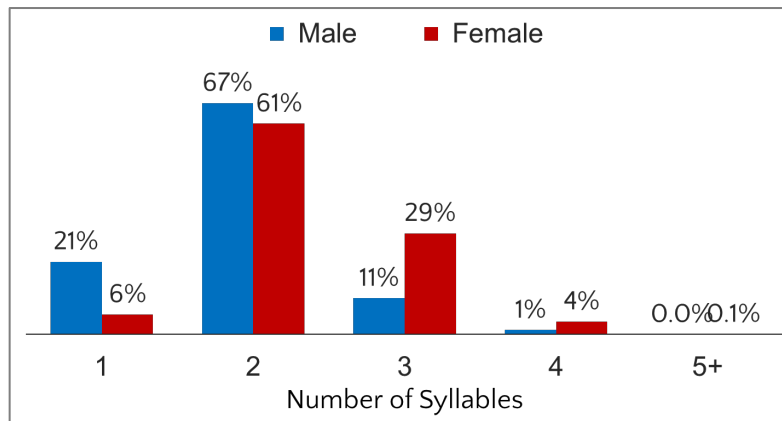


Figure 7. Distribution (out of 100%) of male and female population relative to names syllable count, 1880-2017

**4. Discussion.** Some limitations for this study came from the dataset itself. As the US Social Security Act and Administration was established in 1935, it was only after this year that data started to be collected on all newborns in the US. Thus, the data before 1935 is incomplete relative to the newborn population then. The data also does not include information on race or immigration history to analyze culture or immigration impact on American names.

Building on this study, there are several areas in which there could be further research. First, as this study focused on American names, it would be interesting to apply the same methodology to names from another culture and understand if similar phonological patterns exist between the male and female names of that culture. Second, in addition to names of people, this research could be expanded to study brand and product names. The social perception of brand names can be critical to the success of the underlying businesses, and understanding the potential perceptions will be useful for naming new brands and products, which can then be customized to generate better perception with the target audience.

**5. Conclusion.** Through analyzing 348 million American names over 137 years, this study identified the naming trends from sociolinguistic impacts: the rise of individualism, unisex names, the higher number of unique female names, and the impact of public figures and pop culture. By adding IPA to the names and analyzing the names phonologically, the study showed the significant contrast between male and female names, which suggested an inherent psychological correlation between names' phonology and perception of genders. On average, female names were found to have 25% more vowels than males, and 37% of all female names in the dataset ended in vowels while only 8% of all male names ended in vowels. In addition, there were certain vowels that, when word-final, were linked mostly to female names. Conversely, there were certain consonants that, when word-final, were linked mostly to male names. Male and female names also differ in syllable count – proportionally more female names have three syllables and more male names have one syllable.

While this study focused on American names, the approach and methodology can be applied to names from other cultures which are expected to have different patterns that drive their own social perceptions and naming trends.

## References

- Barry, Herbert Jr. & Aylene S. Harper. 1995. Increased choice of female phonetic attributes in first names. *Sex Roles* 32. 809–819. <https://doi.org/10.1007/BF01560190>.
- Bertrand, Marianne & Sendhil Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review* 4(4). 991–1013. <https://doi.org/10.1257/0002828042002561>.
- The CMU Pronouncing Dictionary. Retrieved August 15, 2018, from <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.
- Hahn, Matthew W. & R. Alexander Bentley. 2003. Drift as a mechanism for cultural change: An example from baby names. *Proceedings of the Royal Society of London B: Biological Sciences* 270. <https://doi.org/10.1098/rsbl.2003.0045>.
- LOGIOS Lexicon Tool. Retrieved August 15, 2018, from <http://www.speech.cs.cmu.edu/tools/lextool.html>.
- Nick, I. M. 2017. Names, grades, and metamorphosis: A small-scale socio-onomastic investigation into the effects of ethnicity and gender-marked personal names on the pedagogical assessments of a grade school essay. *Names: A Journal of Onomastics* 65(3). 129–142. <https://doi.org/10.1080/00277738.2017.1304100>.
- Pelham, Brett W., Matthew C. Mirenberg & John T. Jones. 2002. Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology* 82(4). 469–487. <https://doi.org/10.1037/0022-3514.82.4.469>.
- Social Security Administration, Beyond the Top 1000 Names. Retrieved December 26, 2018 from <https://www.ssa.gov/oact/babynames/limits.html>.
- Twenge, Jean M., Emodish M. Abebe & W. Keith Campbell. 2010. Fitting in or standing out: Trends in American parents' choices for children's names, 1880–2007. *Social Psychological and Personality Science* 1(1), 19–25. <https://doi.org/10.1177/1948550609349515>.
- Zwebner, Yonat, Anne-Laure Sellier, Nir Rosenfeld, Jacob Goldenberg & Ruth Mayo. 2017. We look like our names: The manifestation of name stereotypes in facial appearance. *Journal of Personality and Social Psychology* 112(4). 527–554. <http://dx.doi.org/10.1037/pspa0000076>.