



## Evolution of popularity in given names



Mi Jin Lee<sup>a</sup>, Woo Seong Jo<sup>a</sup>, Il Gu Yi<sup>a</sup>, Seung Ki Baek<sup>b,\*</sup>, Beom Jun Kim<sup>a,\*</sup>

<sup>a</sup> Department of Physics, Sungkyunkwan University, Suwon 440-746, Republic of Korea

<sup>b</sup> Department of Physics, Pukyong National University, Busan 608-737, Republic of Korea

### HIGHLIGHTS

- We find that family and given names play complementary roles to each other for the identification of individuals in a society.
- A pattern of rise-and-fall of popularity for top 100 given names in several countries is observed with the time scale of one generation.
- A diversity of given names is measured as a function of time. It reflects the sociocultural changes in the past.

### ARTICLE INFO

#### Article history:

Received 29 April 2015

Available online 3 October 2015

#### Keywords:

Given names

Popularity

Diversity

### ABSTRACT

An individual's identity in a human society is specified by his or her name. Differently from family names, usually inherited from fathers, a given name for a child is often chosen at the parents' disposal. However, their decision cannot be made in a vacuum but affected by social conventions and trends. Furthermore, such social pressure changes in time, as new names gain popularity while some other names are gradually forgotten. In this paper, we investigate how popularity of given names has evolved over the last century by using datasets collected in Korea, the province of Quebec in Canada, and the United States. In each of these countries, the average popularity of given names exhibits typical patterns of rise and fall with a time scale of about one generation. We also observe that notable changes of diversity in given names signal major social changes.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Since Galton statistically investigated extinction of families, many researchers have studied dynamics of family names [1–7]. The dynamics is well suited to mathematical analysis, because family names are paternally inherited like the Y chromosome in most cases (see, e.g., Ref. [8] for a review). If we look at statistics of family names, the rank–size distribution is broad in many countries [1,2], whereas a clear exponential form is observed in Korea [3–5]. These statistics can readily be explained by the branching process in mathematics, and the essential ingredient to explain the Korean case turns out to be a social taboo on changing family names [3]. Although we have good mathematical understanding on its origin, the exponential rank–size distribution in Korea actually raises another question: It has a characteristic rank scale beyond which minor family names are found with very small frequencies. Indeed, the top ten family names occupy roughly two thirds of the total Korean population, which implies that it is virtually impossible to identify individuals by using family names. Then, how do they distinguish two different persons? An obvious answer would be that the distinguishability is supplied by given names,<sup>1</sup> and one of our goals in this work is to examine whether this statement is justified empirically. Differently from

\* Corresponding authors.

E-mail addresses: [seungki@pknu.ac.kr](mailto:seungki@pknu.ac.kr) (S.K. Baek), [beomjun@skku.edu](mailto:beomjun@skku.edu) (B.J. Kim).

<sup>1</sup> It does not mean that Koreans use a first-name basis in the daily life. Rather, they have taboos on mentioning elders' given names.

family names, parents have a broad spectrum of possible choices in picking up a given name, and the only criterion is that it sounds good and proper. This is, however, rather subjective, and what is worse is that the criterion itself changes from generation to generation, and from one place to the other. Nevertheless, researchers have tried to understand the given-name dynamics by using empirical data [9–12], and a recent study suggests a typical temporal pattern of rise and fall [13]. However, the suggested pattern heavily relies on a simplified model and no information is provided on its characteristic time scale. Therefore, we will verify the existence of such a pattern and estimate its time scale on an empirical basis.

The rank–size distribution and temporal dynamics together determine the diversity of names. If this is directly related to distinguishability as argued above, it will vary with the typical radius of social interactions: For example, if one can live the whole life in a community consisting of a small number of people, we would not need so many names. In a modern society, however, the range of social interactions can be very large, and it is no longer possible to define an individual in ‘relative’ coordinates like someone’s son or someone’s mother. Such a modern society is sometimes called anonymous, but it is actually in this situation that we expect the maximal diversity of names to distinguish every different person. We note that the social status of Korean women has drastically changed, with increasing the radius of social interactions, over the last century. Therefore, we hypothesize that Korean female names have gradually become more and more diverse, compared to the cases of the other countries, which will also be checked in this work.

The present paper is organized as follows: We introduce the datasets used in this study in Section 2. Two main results of the data analysis, i.e., the rank–size distribution and the temporal pattern of popularity, are described in Section 3.1 and Section 3.2, respectively. In Section 3.3, we present how the diversity of given names in each dataset has evolved over the 20th century and discuss major changes in diversity. We then summarize this work in Section 4.

## 2. Datasets

This work analyzes 10 family books in Korea, and some of us have already used them in previous works [3,5,14]. From these datasets, we extract the daughters and women married into the families, and obtain their names and years of birth. Although the family books cover several centuries, we obtain a significant number of female names only for the 20th century. Even in the early 20th century, it was not uncommon for a girl to have no particular given name. To have more female names in our dataset, we additionally include a list of female students enrolled in a university in Korea during 1926–1985. This comprises about 6% of the number of individuals in our Korean female dataset. On the other hand, we exclude male names in the family books from our analysis, because they are affected too much by a cultural constraint: Most of Korean male names consist of two syllables and one of them is often shared by all the male cousins. For example, one of the authors of the present paper has a name consisting of ‘Beom’ and ‘Jun’, and his two other brothers Han Jun and Seong Jun share the latter syllable with him. In this sense, we may regard only ‘Beom’ as his true identifier, whereas ‘Jun’ is an index for the generation in the Kim family. Although the total number of male names in the ten family books is not small, we still find it doubtful that the statistics is enough to neutralize such a distortion. In contrast, the brides and the students are sampled from the whole population and it is hard to imagine any preference for their names. We thus believe that the set of female names in our data can be regarded as an unbiased sample of the whole female population in the past.

For comparison, we will also use a dataset of Quebec in Canada [15,16], in which the most popular 275 female and 200 male names have been recorded annually with their frequencies. In addition, we use another dataset of the United States (US), which includes all the names that are given to more than five newborn babies every year [17]. Due to the one-hundred year time span of the Korean data, we consider the same period from 1900 to 2000 for all the others as well. The number of persons in each dataset is listed as follows: 342,370 females in Korea, 1,203,575 females and 1,205,453 males in Quebec, and 163,523,372 females and 166,237,403 males in the US, respectively.

## 3. Results

### 3.1. Rank–size distribution

We first study rank–size distribution of given names for each dataset. In most countries except a few, the rank–size distribution for *family names* is described as a power law [3,8]. The broadness indicates that it is usually enough to use family names for distinguishing someone from the others. The situation is completely different for Korean family names, so they have an expression roughly translated as ‘going to Seoul to find someone called Mr. Kim’, which basically means ‘a needle in a haystack’. The consideration above naturally leads us to the following idea. In any human society, the combination of family and given names will have resolution to distinguish one individual from the others. If family names already have broad rank–size distribution as in Western countries, given names do not have to provide further distinguishability. On the other hand, if family names have narrow rank–size distribution as found in Korea, given names must be broadly distributed to make every individual identifiable. In other words, we expect the role of given names to be complementary to that of family names.

Let the size  $S$  of a name denote the total number of individuals given the name during the 20th century. We assign a rank  $r$  to each given name after sorting the data in descending order of  $S$  (Table 1). By construction,  $S(r)$  is a non-increasing function. We indeed find that the rank–size distribution of Korean female names can be fitted to a power-law form as depicted in

Download English Version:

<https://daneshyari.com/en/article/7379043>

Download Persian Version:

<https://daneshyari.com/article/7379043>

[Daneshyari.com](https://daneshyari.com)