Large-scale statistical analysis on representation of public figures in newspapers and naming conventions in Finland in 1900–1939

Antti Kanner & Jaakko Raunamaa

Abstract: This article examines the influence public figures had on the name-giving in Finland from 1900 to 1939 via their presence in newspapers. The study is based on digital materials and the results were obtained using statistical research methods. The main finding of the study is that generally, the prevalence of notable peoples' names does not reflect the person in the way people choose first names for their children. On the other hand, more in-depth analyses reveal consistent interaction: short-term peaks in prominence of public figures in the newspaper data coincide with peaks in the popularity of their first names 50% more often than could be expected by random. Also, high-correlation outliers reveal personalities with a definite impact on first given name trends, interpretable as rule-confirming exceptions of the general pattern of very modest correlations.

Keywords: name-giving, statistical analysis, early 20th century, history of Finland, digital humanities.

1. Introduction

This article focuses how newspapers and public figures influenced trends for first names in Finland during the first half of the 20th century (1900–1939). It seeks to answer the following research question: Is there a statistically observable interaction between the prominence of public figures in the newspaper data and trends with first names attributable to an influence of the former to the latter?

The role of external factors (such as popular culture and literature) in name giving is a familiar and discussed topic in onomastics (e.g. Ainiala et al. 2012:178; Leibring 2016; Nyström 2016:12). There are a few examples in Finnish onomastic literature that are often used to demonstrate the influence of public figures and literature characters on name-giving. For example, it is claimed that the popularity of the female name *Armi* increased greatly after 1952 when Armi Kuusela won the Miss Universe pageant (e.g. Kiviniemi 2006:217). In similar fashion, the case of *Olavi* is often used to illustrate the impact of literature on Finnish naming conventions (e.g. Vilkuna et al. 1988:125; Kiviniemi 2006:183).

The most relevant study for this research was conducted by Eero Kiviniemi in 1982 and consisted of surveys organized in the late 1970s that focused on the principles of name giving in Finland (1982:129). The results include close to 600 answers, which, in turn, contain details of approximately 4,500 first names. Based on the outcomes, Kiviniemi created a nine-part classification model to demonstrate the different categories of naming principles and their relative proportions. According to Kiviniemi, three percent of the first names were given based on famous individuals (173). Furthermore, five percent of the cases fell into the category of 'historical and literature characters' (ibid).

Hence, it seems plausible to assume that external cultural and social factors such as celebrity influence name-giving. On the other hand, one must remember that this assumption is based on very limited data. The examples discussed in onomastic literature and Kiviniemi's statistical analysis (1982) only cover a very small portion of all the names given annually in Finland. According to the records of Official

Statistics of Finland (OSF), the average annual number of new-borns at the beginning of the 20th century (1900–1939) was approximately 80,000. The conclusion is that no scholar in Finland (or to the best of our knowledge, nowhere in the world) has comprehensively analysed the influence of external factors in name giving through statistical methods.

Limited access to suitable materials and methods has made it difficult to conduct statistical research on external factors in name-giving from previous decades. Since the 2010s, various archives have been digitized and made available for digital analyses (Ogilvie 2016). It is also clear that the field of digital humanities and the methods used within it have evolved and expanded during the past decades (Brennan 2018).

In the present article, we bring digitized archive materials and robust statistical analysis methods together to investigate the interaction between the occurrences of public figures in Finnish newspapers and the naming conventions in Finland during the first part of the 20th century (1900–1939). More specifically, we have collected a list of names of famous Finns from the National Biography of Finland (NBF) and recorded how often they are mentioned in the Finnish newspapers digitized by the National Library of Finland. This is done in three parts: First, by using Pearson's correlation coefficient to investigate the overall correlation between the occurrences of public figures in the press and the popularity of their first names according to the first given name statistics of Finland.[1] Second, we analyse the predictability of the distribution across all given first names for each year, given the distribution across relative frequencies of public figures' mentions for the respective years. Third, a peak detection method has been used to detect those names and famous persons whose popularity and prominence have peaked at the same time. Based on the

---

[1] When discussing public figures, the expression *first name* refers to the name by which the person was widely known. It was not necessarily always the person's official first given name. It was common to have two first names in Finland at the beginning of the 20th century (Ainiala et al. 2012:174). See Section 2.2 for more details regarding the choice of names used for public figures.

results, some names are analysed at the more fine-grained level. We discuss the reasons and features that have either enabled or prevented the popularity of a first name being influenced by the prominence of a public figure.

The article comprises four sections. The first one provides an overview of Finnish history from the latter decades of the 19th century to the end of the 1930s. The second section focuses on the research materials and more specifically explains how the data is compiled and adjusted for the purposes of this work. Issues with the research material are also indicated and discussed. The analyses and their results are presented in the third section. We have selected three complementary statistical analyses for our study and each of these will be presented alongside their results. This choice of structure keeps the methodological description of each analysis in close proximity to its results. Thus, it is more reader-friendly for audiences who are not well-versed in statistical details.

Because large-scale statistical analyses have not previously been conducted on our material, we chose to apply the most intuitive and robust statistical methods for looking at the interaction between press prominence and the popularity of first names. The fourth and final part presents the conclusions together with two more detailed analyses of public figures and their influence on how names are chosen.

We also argue that while during the period of our study celebrity culture, newspaper media and naming conventions all were quite different to what they are today, the interaction of the three can be meaningfully studied as long as interpretations of the results do not rely on unfounded assumptions regarding their historical nature. For this reason, we will seek to present our results in robust statistical terms and only offer reserved interpretations of the changes in underlying cultural, social and linguistic structures involved. We see that our study can form a starting point for more specific research. Future contributions employing more elaborated and detailed statistical analyses may then benefit from the groundwork laid out here.

2. Historical background

There are many examples of major changes in naming conventions coinciding with the most important social and cultural developments in the surrounding society (Salway 1994; Raunamaa 2021:72). This line of thought is further supported by early 20th century Finland. To begin with, naming choices changed during the first part of the 1900s. From the Late Middle Ages onwards (c. 1400 CE, most of the inhabitants of Finland were named after Christian saints and their names recorded mainly in their Swedish forms (Ainiala et al. 2012:161–166; Raunamaa 2020). The system, which had remained fairly stagnant for centuries, began to break down from the end of the 19th century. In short, the popularity of Finnish names started to rise, whereas the number of names written in non-Finnish form declined rapidly. This phenomenon is depicted in Figure 1 below (based on Kiviniemi 2006:265). One can notice how the rapidly declining popularity of the first category "Names in foreign form" (e.g. *Anders*, *Jakob* and *Margareta*) coincides with rise of second and third sections that include Fennicised foreign names (= names adapted into Finnish) (e.g. *Anders -> Antti*, *Jakob -> Jaakko* and *Margareta -> Reetta*) and names based on pre-Christian Finnish nomenclature (e.g. *Kauko*), Finnish mythology (e.g. *Väinö*) or Finnish vocabulary (e.g. *Urho*). These changes in naming conventions concurred with turmoil in Finnish society. The events leading to the social upheavals and concurrently created a fertile ground for changes in the use of given names in Finland at the turn of the 20th century are described in more detail below.[2]

[2] This section is based on Saari (2012) and Ylikangas (2002:103–189) unless stated otherwise.
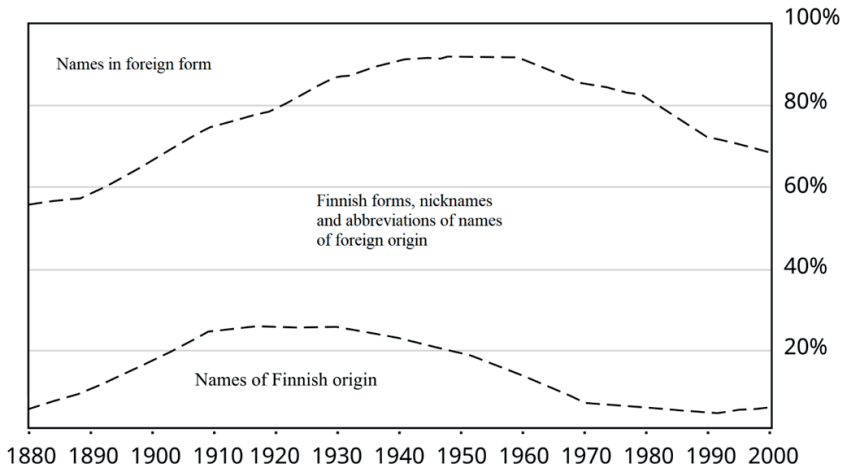
Figure 1. Proportions of different groups within the names given in Finland between 1880 and 2000 based on Kiviniemi 2006: 265. The upper section "Names in foreign form" refers to names that were used in their original foreign form and were not Fennicised. The second section "Finnish forms, nicknames and abbreviations of names of foreign origin" includes all the foreign originated names that were adapted into Finnish. The third category, "Names of Finnish origin", covers all the names that derive from pre-Christian Finnish nomenclature, Finnish mythology or were created on the basis of Finnish language from the mid-19th century onwards.

During the latter part of the 19th century, the Grand Duchy of Finland was linguistically a diglossic society: Swedish was predominantly used as the prestige language of administration, culture, politics, public sphere and education while Finnish was the household language of rural population (Marjanen et al. 2019:57). However, towards the turn of the century, Finnish gradually became more prominent in public discourse and administrative use and especially as the symbol of national identity. Crucial steps for the wide adoption of Finnish language were, for example, the establishment of the first Finnish teacher-training college in 1863 and Finnish achieving equivalent legal status to Swedish in 1902 (Saari 2012:180). Finnish education institutions also developed rapidly during the 19th century. At the beginning of the 1800s many Finns had basic literacy as the Lutheran Church required that adults could read religious texts (Kuikka 2010:2). From

1850 onwards, the national government became interested in the education of ordinary people (ibid. 5–9). The result was that by the turn of the 20th century, most of the Finns were able to read.

The improving prestige of the Finnish language was influenced by the European intellectual trends of nationalism and romanticism. When these ideological trends arrived in Finland, the local cultural elite, mostly Swedish speaking at the time, started to promote a distinct national identity of Finns, drawing clear separations from the neighbouring nations, especially Russians and Swedes. Improving the status of Finnish language was seen as one of the key expressions of the national identity and formed the core of the political agenda of the so-called Fennoman political movement.

The Russian Empire showed mixed reactions to Finnish nationalistic developments. On the one hand, Finnish national identity was seen as instrumental in severing the ties Finns still had to the Swedish Realm under which Finland had belonged before the Finnish War (1808–1809). On the other hand, a distinct Finnish national identity also hindered the assimilation of Finns as loyal subjects of the Russian Empire. During the 19th century, the position of the Russian government shifted between these two emphases, the shifts resulting in legislation sometimes promoting and sometimes demoting the status of Finnish as an official language. From the perspective of this article, one of the most impactful periods started at the turn of the 20th century when Russification started with the aim to linguistically, culturally and economically absorb Finland into Russia. This period lasted until independence. One of the consequences of Russification was that Finns (both Finnish and Swedish speaking) united against Russification and the trend of nationalism was strengthened.

Besides linguistic and cultural turmoil, it is also safe to claim that Finnish society also underwent radical economic and social changes. At the beginning of the Russian era, Finland was still an agrarian society, but from the 1830s onwards industrialization took hold and the economy grew. Paper was one of the key goods produced in Finland, which, of course, contributed to the development of the local newspaper publishing.

In 1917, after many years of unrest and upheaval within the Russian Empire, the Parliament of Finland declared independence. The declaration was soon followed by political turmoil, which quickly transformed into an open conflict. The result was a civil war between the Conservative 'White' and Revolutionary Communist 'Red' armies, the former representing the forces of Finnish Socialist Workers' Republic and the latter being a coalition of various groups united by the opposition of the communist and left-wing Reds.

The Whites won the short but fierce war, after which society rapidly began to normalize, as did economic growth. This was supported in particular by the forest industry as well as the economic openness that Finland promoted after the civil war. The new nation also craved fresh national heroes which were obtained, for example, among the long-distance runners and Nordic skiers who brought many Olympic gold medals to Finland.

3. Research material

This section gives a description of the materials used in this study. We start with the newspapers, continue with public figures and finally, describe the material containing the Finnish first name statistics. We also discuss the issues related to the materials in the subsections.

3.1 Newspaper material

Newspapers were the most widespread form of media in the early 20th century Finland, making them the most viable source for investigating the influence of public figures on name-giving in Finland. Since most of the Finns were introduced to newspapers by the end of 19th century (Marjanen et al. 2019:56), the degree of literacy even in rural regions of Finland was relatively high (Suomen tilastollinen vuosikirja, 1903) and as radio did not become widespread until the 1930s (Endén et al. 1996:67), the printed press was likely the main source of information at the time. However, it must be taken into con-

sideration that, although newspapers were widely published during the early 20th century, regional differences were still significant. For example, in 1920 many rural areas were hundreds of kilometres away from the closest newspaper publisher and therefore, it is likely that the inhabitants of these remote regions did not encounter newspapers regularly (Marjanen et al. 2019:71).

Second, the historical newspapers in Finland are readily available for research. Many of the early 20th century Finnish newspapers have been digitized by the National Library of Finland and their accessibility varies across several platforms. The data we have used come from the annotated corpus version maintained by the Language Bank of Finland (www.kielipankki.fi). It was extracted using the API interface of the Korp corpus tool by which Kielipankki provides access to their resources (Borin et al. 2012). We only included Finnish language newspapers published between 1900–1939. From these newspapers, we collected yearly relative frequency counts for all sequences matching the full names in the Finnish National Biography.

Due to licensing issues, the Kielipankki version of the historical newspaper collection only has all papers until 1910. After that, the selection of newspapers is much smaller: in 1910 the number of tokens is close to 261 million, whereas the 1920 data contain approximately 20 million tokens. Although using the relative frequencies mitigates the immediate issues related to uneven sizes of the yearly corpora, the problems related to unpredictable sampling of the later years remain. The fact of the matter is that we cannot know whether the latter two thirds of our data are systematically biased compared to the first third. This can only be validated in the future once the rest of the material becomes available for research. For now, we simply assume that the variation in the inclusion of newspapers in the corpus (i.e. exclusion of most of the newspapers from the later yearly corpora) does not systematically affect the prominence of certain public figures and present our results as conditional to that assumption.

A summary inspection of some public figures and the relative frequencies in the newspaper corpus does not seem to depend on the differences in the composition of the corpus, or at least to a great extent.

Below (Figure 2) are relative frequency graphs for two famous Finns, long distance runner Paavo Nurmi and composer Jean Sibelius, both widely known public Finnish figures of their time. Both graphs show variation in frequency congruent with their respective biographies, but not at least obviously attributable to the major segmental variation in the data. More importantly, the scale of the relative frequencies also seems to be well comparable across the years, with these two figures reaching peak prominence in around one to two occurrences per million tokens of running text across the period.
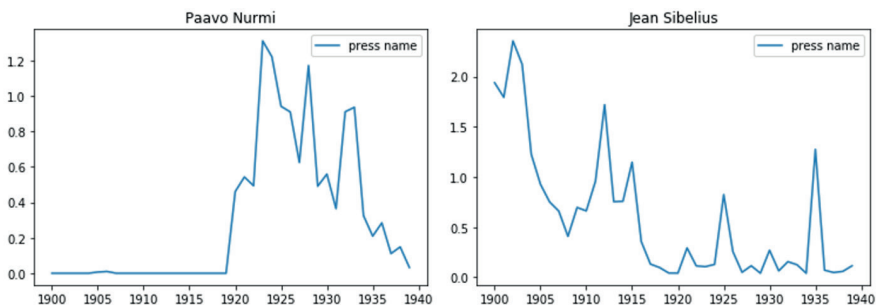

Figure 2. Relative frequency graphs of occurrences in the newspapers for two famous Finns, long distance runner Paavo Nurmi and composer Jean Sibelius.

3.2 List of public figures

A list of those public figures who were active in Finland during 1900–1939 is derived from the National Biography of Finland (NBF). In total, this material contains more than 6,500 Finnish biographies. The backgrounds of the selected individuals vary, but the most common categories of those who were active during the 20th century are (starting with the largest group) members of the government, visual arts & photography, literature, humanities and military & national defence. Our material is based on a dataset created by the Biografiasampo project (BS) (Hyvönen et al. 2018) and obtained through Yasgui SPARQL query interface. Our query was searching for public figures born between 1850 and 1939. The outcome contains 3,393

public figures. We removed all those figures that had three or less first name occurrences in our data but made no other frequency-based filtering. After this procedure, we comprised a list of 2,649 public figures, including information on their sex, date of birth and class categories. The majority of these notables (2,291) are men and the minority (358) are women.

The main issue with the above presented list of public figures is that it is not complete. There are probably many such notables that were often mentioned in the press but were not considered important enough to be added to the National Biography. Another problem is that the material mainly contains Finns and accordingly, many well-known foreign public figures are excluded. The names of the public figures are an issue as well. Especially at the beginning of the 20th century, there was variation in the spelling of names, and many notables were not presented under their official names. This is particularly true for those public figures who were Finnish speakers but were baptised with a Swedish name. However, it seems that at least the majority of public figures are in the BS/NBF dataset with the name they were known for.[3] Furthermore, there are some public figures whose prominence is exaggerated as their names occurred in company names as well (e.g. *Werner Söderström* and *Victor Ek*).

Despite the above-described issues, the number of public figures derived from NBF is extensive and the list consists of people from a wide range of backgrounds. Furthermore, there is extensive press coverage of the majority of the public figures. Thus, it can be argued that the public figures listed in NBF are a comprehensive sample of those who could have influenced Finnish naming conventions at the beginning of the 20th century.

---

[3] For example, famous Finnish sportsmen *Hannes Kolehmainen* and *Ville Ritola* are listed with the names they were known for although they were officially *Johannes* and *Viljo* (BS/NBF).

3.3 First name statistics of Finland

Statistics of first given names in Finland have been obtained from the records of the Finnish Population Register Centre.[4] The material contains annual details of first given names, their numbers and gender.[5] Our dataset is limited to the years 1900–1939. Furthermore, we obtained Finland's annual birth rate statistics from Statistics Finland and, using these figures, we calculated the relative annual frequency of each name. The data contain 87,406 different names.

Although the records of the Finnish Population Register Centre are comprehensive, it should still be emphasized that they do not always depict the real usage of the names. This is especially true at the beginning of the 20th century when many Finnish speaking new-borns were still baptised with Swedish names although in speech, they were addressed differently. On the other hand, this issue is also an interesting subject for research, as one could assume that the prominences of Finnish and Swedish names in the newspaper data would be in line with the Finnish and Swedish first given names in the official name records.

4. Methods and results

In this section, we present our three computational methods and the results they have yielded. We start with the Pearson's correlation coefficient, then move to Shannon-Jensen Divergence and finally present our analysis regarding coinciding peaks. The results of each are presented immediately following a brief discussion of what the method seeks to capture and what are the major reservations concerning their use.

---

[4] We obtained the name statistics through the GitHub page of the Finnish Broadcasting Company (YLE): https://github.com/Yleisradio/yle-uutiset/tree/master/names.

[5] For privacy reasons, the value 10 is used when the annual frequency of a name is between one and ten.

4.1 Pearson's correlation coefficient

We sought to test the hypothesis of the influence of public figures' degree of prevalence in the newspapers and popularity of their respective first names with as simple, robust and intuitive method as possible. Perhaps the most common approach for an analysis of this kind conceptualizes the interaction between two phenomena as that of mutual dependence and correlation. In other words, the number of occurrences of public figures' names per year is seen as the independent variable and the number of their first names in name records of the same year form the dependent variable. As we aim to provide the most robust statistical observations in the matter, we study the dependence of these two variables as that of linear correlation and select Pearson's correlation coefficient for measuring the degree of correlation between them.

We calculated the Pearson correlation coefficient between the yearly relative frequencies of each of our 2,649 public figures full name mentions in the newspaper data and the relative frequencies of their first name in the name records. This yielded a correlation score for each name, which is represented in the histogram (Figure 3) below:
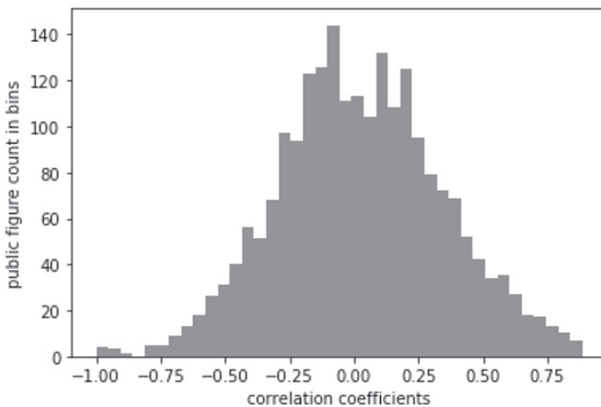

Figure 3. Distribution of correlation coefficients for each NBF public figure as a histogram.

The distribution in Figure 3 is reminiscent of a multimodal distribution with two modes: one near –0.3 and the other near 0.2. While this is surprising at first, this bimodality has a ready explanation. As is noted in Figure 1 in Section 1.1, Finnish first name conventions underwent a relatively remarkable tectonic shift during the period being studied, as names congruent with non-Finnish orthography steadily decreased in both type and token frequency. This means that public figures whose peak of fame gradually declined during the period automatically have some correlation with the frequency of their first name if they happened to have an orthographically non-Finnish name. Similar trends can be seen in name variants with explicitly Finnish orthography: almost all become more frequent over time and thus produce correlation with public figures with growing fame. Thus, a modest positive or negative correlation is to be expected here.

While correlation does not entail causality, it can be reasonably surmised that a very simple and blunt version of causation between public figure's names and first given name trends should result in linear correlation between mentions and given names. As no such correlation is detectable in Figure 3, apart from the modest bimodal peaks explainable by broader patterns of change in names, we can conclude that the frequency of first names is largely independent of frequencies of public figures mentioned in the newspapers. However, Figure 3 also shows possible outliers at both ends of the histogram, where positive or negative correlations are remarkably strong.[6] Some true cases of exceptional mutual co-dependence could indeed be found among these sections, but they do little to challenge the overall interpretation.

4.2 Jensen-Shannon Divergence

However, there is another, complementary, possibility for looking into this matter. While the correlation coefficient focused on the temporal variation of individual names and sought to detect patterns of co-dependency, the repertoires of mentions and given names can be studied

---

[6] The outliers are represented by the bump near –1.0 mark and comparatively heavier mass near 0.75.

as a whole. In this line of analysis, the question is instead formulated as, 'How well does the distribution across all public figure occurrences predict the distribution of first given names for every year?' A common method used for such analyses is the Küllbeck–Leibler Divergence (KLD) (cf. e.g. Degaetano-Ortlieb and Teich 2018). However, as not all the public figures appear in newspapers every year, nor do their first names appear in name records every year, we opted to use its symmetrical variant Jensen–Shannon Divergence (JSD), following an approach established in similar studies using language data (Pechenick et al. 2015). While mathematically alike in many regards, unlike KLD, JSD yields identical divergence scores regardless of which distribution is selected for base of comparison. The values of JSD range from 0 (meaning the distributions are the same or identical) to 1 (meaning they are maximally different).

The prevalence distribution was obtained by summing over relative frequencies of occurrences of each public figure with the same first given name each year. That distribution was then compared to the distribution of first given names for the same year by computing the JSD similarity between them. Because the data consist of potential confounding factors related to wider patterns of change in naming practices, as was noted above, the JSD similarity scores were compared to a confidence interval obtained by a bootstrapping process. In the bootstrapping, each yearly mention distribution was compared to given first name distributions for every year. The maximum, minimum and mean of these 28 scores were noted and plotted in Figure 4 below. If there was a systematic correspondence between distributions over mentions and given first names, it would be fair to assume that this correspondence was strongest between distributions obtained from the same year rather than random pairings of year.
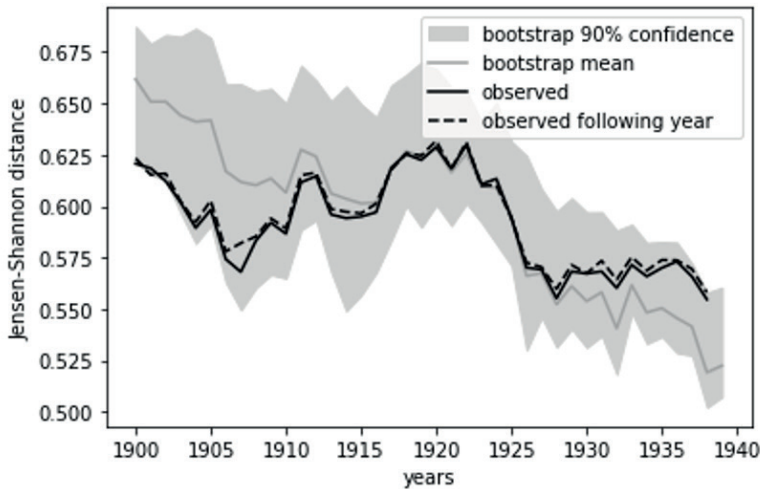
Figure 4. Jensen-Shannon Divergences of consecutive years. The grey area marks the lowest and highest scores when each year is compared to every other year. The grey line is the mean score across all years and black line is the score for the same year.

No such correspondence is visible in Figure 4 above, perhaps apart from the first five years. Rather, generally, the JSD score between the distributions of the same year are neither stronger nor weaker than could be expected when comparing distributions from two random years. This observation complements the one in Figure 3, by giving further grounds to reject an overarching hypotheses of interaction between the occurrences of public figures in the newspapers and their first names in name records. The exceptional pattern in the first five years does however warrant some further questions on whether that truly would be a sign of better predictive power of prevalence distributions. On the other hand, it might simply be to do with the fact that as some (mainly Swedish, see Section 1.1 for more details regarding this development) names present in the data at that time fall out of use. For that reason alone, the earliest years are more similar than the later years. This should result in both the early end having high similarity (low JSD score) and the latter end having low similarity (high JSD scores), which is congruent with Figure 3 above. For the most

part, it can be noted that there is no clear evidence supporting a strong correspondence between mentions and given names.

4.3 Coinciding peaks

The analyses above sought to look at very broad trends and temporal variation and investigate whether changes in degrees of publicity resulted in names becoming more frequent or infrequent over time. These resulted in largely negative results. However, previous studies indicate another pattern of interaction between degree of public presence and naming conventions; Kiviniemi (1982:173) observed that questionnaire data showed how roughly 3% of parents were inspired by public figures when selecting first given names for their children. In addition to this survey study, Kiviniemi (e.g. 2006:212–226) has presented many examples of cases where publicity has influenced the popularity of a name.

Examining peaks in popularity of first given names was our third approach to studying the influence of public figures on naming conventions. Here, we calculated how many of the peaks in prevalence graphs coincided with a peak in the first given name graph. Because the prevalence time series tend to demonstrate distinctive spikes, we again compared the observed proportions to a confidence interval obtained by bootstrapping. This time, the 90% bootstrap interval was based on comparing each public figure with 1,000 random first names and selecting the 0.05...0.95 of that distribution.

Also, because there is no single, straightforward definition for a peak, many different range-based definitions were studied. In other words, peaks were defined as the highest value in running ranges of 2 to 19 years, resulting in 18 different definitions. The results of these analyses are presented in Figure 5 below.
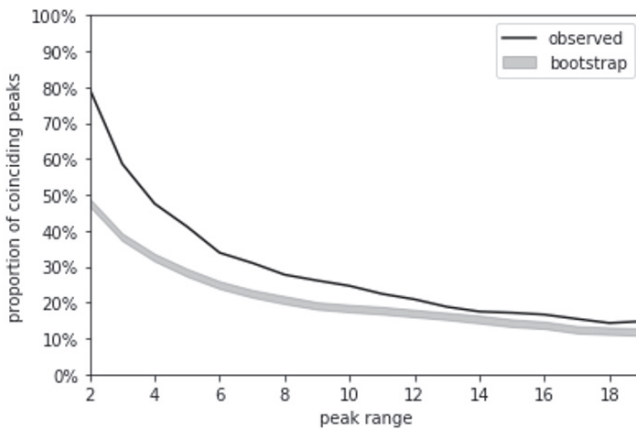
Figure 5. The proportion of coinciding peaks in newspaper occurrences and first given names in name record data. The grey area is 90% bootstrap confidence interval for 1,000 random public figures – first name pairings.

The vertical axis shows the proportions of coinciding peaks, the horizontal axis the different peak definitions running from 2 to 19. The black line represents the actual observed proportions for coinciding peaks between the public figure occurrences and their respective first names. The grey area represents the randomly obtained 90% confidence interval. As can be seen in Figure 5, the observed line stays well above the confidence interval with diminishing difference between them. Even until the peak range of ten years, there are roughly 10% additional coinciding peaks above what could be expected by chance alone. This not only corroborates Kiviniemi's observation (1982:173) but indicates that his estimation of 3% might even be conservative. The interpretation here would be that over the span of ten years, public figures' names would cause three-times as many peaks in popularity of names in first given name records compared to questionnaire data.

5. Conclusions and discussion

Our aim was to use computational methods to investigate what role public figures have had in determining how first given names were chosen in Finland at the beginning of the 20th century. As far as we are aware, this kind of research has not been conducted before. Thus, one of our main focuses was learning if computational research methods can be used to conduct such a study.

The results corroborated many of the earlier observations made from smaller scale studies. There seems to be no over-encompassing influence of celebrity names influencing how people choose to name their children, or that are at least clear enough to register on such metrics as general correlation or Jensen-Shannon Divergence. It should be noted, however, that metrics such as correlation coefficients and JSD give more emphasis on how the overall probability mass is divided across the distributions than matching peaks, and so have limited interpretability concerning possible causal projections. However, the results regarding the material available is provisional: the number of occurrences of public figures in the newspaper data could possibly change drastically once all of the material becomes available. Nevertheless, the current results seem to give reason to assume this will be relatively unlikely, since there seems to be quite few dramatic changes in distributions that would correspond with different compositions of the dataset.

Furthermore, it is possible that selecting only specific subgroups of public figures (such as prominent cultural figures) might show different levels of correlation. Additionally, it seems that even names with relatively high levels of correlation did not match the given name statistics close enough to suggest causation.

A closer look at individual names shows that often, the influence of public figures on the naming conventions has been much more nuanced than can be studied using statistical methods alone. As noted earlier, part of the positive overall correlation is caused due the decline in foreign names (mainly Swedish ones) at the beginning of the 20th century, both in newspapers and in given names. Needless to say, this phenomenon was induced not only because of the examples set by the public figures but mostly because of the ongoing Finnicization. As

the proportion of foreign names (see Figure 2) had decreased rapidly from the 1880s onwards, it is to be expected that their number also declined among the local public figures.

In a similar fashion, it is difficult to measure the influence of public figures in those cases where the name was common among the notables and ordinary people. The name *Juhani* is a good example. Its popularity started to increase from the mid-1920s onwards as indicated by the blue graph in Figure 6 below. The same figure also depicts the newspaper prevalence of car merchant *Juhani Korpivaara* (orange). Based on the Pearson coefficient, these two variables have a strong correlation. However, one can easily notice that the popularity of the name *Juhani* had already increased before any mentions of *Juhani Korpivaara* in the newspapers. Furthermore, the list of the most frequent public figures in the press contains many people whose first name was *Juhani*. Among these men, the author *Juhani Aho* received the most mentions. However, at the peak of his press appearances, the name had already been prevalent since the start of the 1910s, therefore it seems unlikely that his example would have been significant in increasing the popularity of the name *Juhani*. The conclusion is that there is no single public figure named *Juhani* who could be credited with initiating the name's popularity.
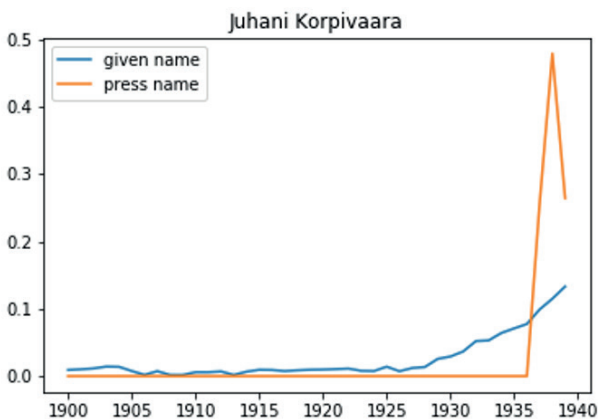

Figure 6. The prevalence frequency of *Juhani Korpivaara* in the press (orange) and the popularity of the name *Juhani* according to the Finnish first given name records (blue).

On the other hand, the degree to which peak numbers of mentions of public figures in the newspaper data matched with peaks of their respective first names in the name records suggests that instead of overarching correlation, the interaction between these two elements is temporally a much more restricted phenomenon. Over a period of ten years, the peaks coincided roughly 50% more often than could be expected by chance. This observation also corresponds with previous studies, where famous people have been seen to generate peaks of popularity of their first names (Kiviniemi 2006:212–226). Based on our results, the earlier assumptions regarding the role of public figures in naming trends can be further elaborated by claiming that this phenomenon was not restricted to only the most famous individuals. In addition to this, it can be argued that often the influence of public figures caused relatively small and temporally limited popularity peaks.

The case of a male actor and director *Kosti Elo* (Figure 7) is a good example of a situation in which press prominence seems to have caused slight rises in the popularity of a name. The correlation appears to be very strong on multiple occasions. For example, during the first years of the 20th century, the rising press prominence of Kosti Elo is followed by an increase in the popularity of his first name *Kosti*. However, the situation is not as clear as it seems. The problem with names like *Kosti* is that their popularity as a first given name was low. The popularity of *Kosti* peaked in 1904, when 78 were given the name. The lowest point was in 1920, when the name was given to 20 people. With such small numbers, the role of chance must also be considered. On the other hand, as pointed out in Section 3.2, the peaks of press prominence and first given name popularity coincide more frequently than could be expected by chance. Thus, it is completely plausible that the relatively small and temporally limited increases in the popularity of a name, like in the case of *Kosti*, can be explained by the press prominence.
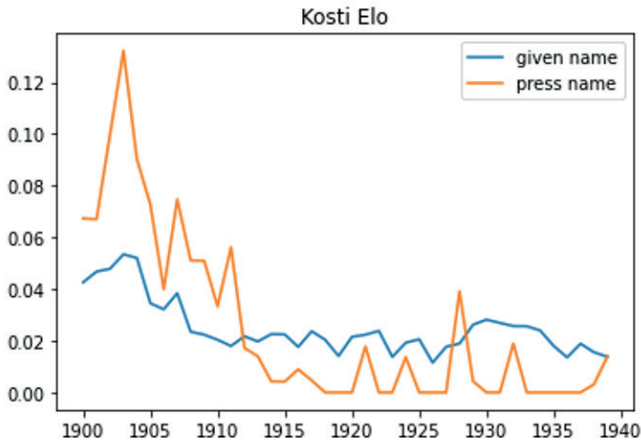
Figure 7. The prevalence frequency of *actor and director* Kosti Elo in the press (orange) and the popularity of the name *Kosti* according to the Finnish first given name records (blue).
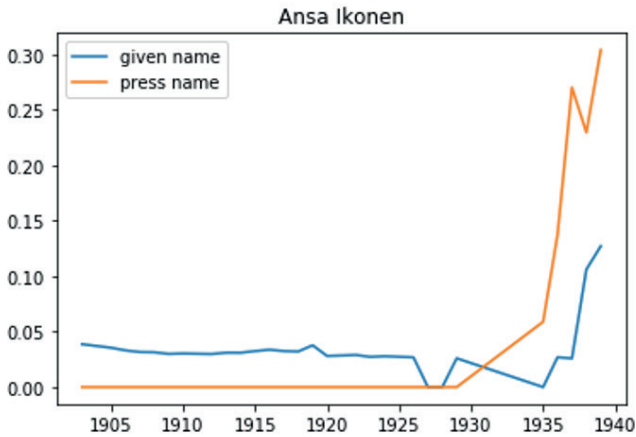

Figure 8. The prevalence frequency of actor *Ansa Ikonen* in the press (orange) and the popularity of the name *Ansa* according to the Finnish first given name records (blue).

There are also cases where the correlation is obvious and strong upon closer inspection. For example, Figure 7, shows how the famous Finnish actor *Ansa Ikonen* and her prevalence in the press (orange) is clearly connected to the growing popularity of the name *Ansa* (blue). We can see that the name *Ansa* was almost non-existent until 1935 and 1936 when Ansa Ikonen became one of the most famous actors in Finland.[7]

The above presented case of *Ansa Ikonen* seems to be a typical example of a situation where the press publicity of a public figure has influenced the naming conventions. What these cases have in common is the low starting level of popularity of the name. As previously stated, more common names like *Juhani* had many other possible factors causing the increase in the popularity. Preliminarily, it could also be argued that female names were more prone to the influences of publicity than the male equivalents. This idea would be in line with earlier notions according to which female name trends tend to vary faster than the male ones (Raunamaa 2020). However, this preliminary notion requires further investigation. The same can be said of the categories of the public figures: which types of public figures influenced naming conventions the most.

In the future, a more elaborate statistical analysis could be developed to take into account the general trend of each name and weighted influence of public figures against that. At the same time, a number of explanatory variables could be considered, such as the gender or professions of public figures. For example, multinomial regression could provide a statistical framework for such an analysis.

References

Ainiala, Terhi, Saarelma, Minna & Sjöblom, Paula. 2016. *Names in focus: An introduction to Finnish onomastics*. Helsinki: Finnish Literature Society.

BS = Biografiasampo. https://biografiasampo.fi/. See project homepage

---

[7] Her real first given name was *Aili*.

https://seco.cs.aalto.fi/projects/biografiasampo/en/ for further information (accessed 20 June 2022).

Borin, Lars & Forsberg, Markus & Roxendal, Johan. 2012. Korp – the corpus infrastructure of Språkbanken. In *Proceedings of the eighth international conference on language resources and evaluation* (LREC'12), 474–478. Istanbul: ELRA.

Brennan, Claire. 2018. Digital humanities, digital methods, digital history, and digital outputs: History writing and the digital revolution. *History Compass* 16(10). https://compass.onlinelibrary.wiley.com/doi/abs/10.1111/hic3.12492 (accessed 20 June 2022).

Degaetano-Ortlieb, Stefania & Teich, Elke. 2018. Using relative entropy for detection and analysis of periods of diachronic linguistic change. *Proceedings of the second joint SIGHUM workshop on computational linguistics for cultural heritage, social sciences, humanities and literature*, 22–33. Santa Fe, New Mexico: Association for Computational Linguistics.

Endén, Rauno & Kokkonen, Jüri & Lyytinen, Eino & McAlester, Gerard. 1996. *Yleisradio 1926–1996: A history of broadcasting in Finland*. Helsinki: Finnish Historical Society.

Hyvönen, Eero & Leskinen, Petri & Tamper, Minna & Tuominen, Jouni. 2018. Semantic National Biography of Finland. In Mäkelä, Eetu & Tolonen, Mikko & Tuominen, Jouni (eds.), *Proceedings of the digital humanities in the nordic countries 3rd conference (DHN 2018)*, 372–385. (CEUR Workshop Proceedings 2084). http://ceur-ws.org/Vol-2084/short12.pdf (accessed 20 June 2022).

Kiviniemi, Eero. 1982. *Rakkaan lapsen monet nimet: suomalaisten etunimet ja nimenvalinta*. Espoo: Weilin + Göös.

Kiviniemi, Eero. 2006. *Suomalaisten etunimet*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kuikka, Martti. 2010. Opettajankoulutus eilen, tänään ja huomenna. *Koulu ja menneisyys* 48. 1–18.

Leibring, Katharina. 2016. Given names in European naming systems. In Hough, Carole (ed.), *The Oxford handbook of names and naming*, 199–213. Oxford: Oxford University Press.

Marjanen, Jani, et al. 2019. A national public sphere? Analysing the language, location and form of newspapers in Finland, 1771–1917. *Journal of European Periodical Studies* 4(1). 54–77.

NBF = National Biography of Finland. Helsinki: Finnish Literature Society, 1997–. https://kansallisbiografia.fi/ (accessed 20 June 2022).

Nyström, Staffan. 2016. Names and meaning. In Hough, Carole (ed.), *The Oxford handbook of names and naming*, 39–51. Oxford: Oxford University Press.

OSF = Official Statistics of Finland: Births [e-publication]. Helsinki: Statistics Finland. http://www.stat.fi/til/synt/tau_en.html (accessed 20 June 2022).

Ogilvie, Brian. 2016. Scientific archives in the age of digitization. *Isis* 107(1). 77–85.

Pechenick, Eitan Adam & Danforth, Christopher & Dodds, Peter Sheridan. 2015. Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PloS ONE* 10(10). https://doi.org/10.1371/journal.pone.0137041 (accessed 20 June 2022).

Raunamaa, Jaakko. 2020. Margareta, Katerina and Kristina: Female names in medieval Finland. *Studia anthroponymica Scandinavica* 35. 75–98.

Raunamaa, Jaakko. 2021. *Finnic anthroponymy in the Middle Ages: Perspectives on pre-Christian and early Christian personal names*. Helsinki: University of Helsinki. (Doctoral dissertation.)

Saari, Mirja. 2012. The development of Finnish into a national language. In Hüning, Matthias & Vogl, Ulrike & Moliner, Oliver (eds.), *Standard languages and multilingualism in European history*, 179–204. Amsterdam: Benjamins.

Salway, Benet. 1994. What's in a name? A survey of Roman onomastic practice from c. 700 BC to AD 700. *The Journal of Roman Studies* 84. 124–145.

Suomen tilastollinen vuosikirja 1903 = Suomen tilastollinen vuosikirja 1903. Helsinki: Tilastollinen päätoimisto.

Vilkuna, Kustaa & Mikkonen, Pirjo & Paikkala, Sirkka. 1988. *Uusi suomalainen nimikirja*. Helsinki: Otava.

Ylikangas, Heikki. 2002. *Käännekohdat Suomen historiassa: pohdiskeluja kehityslinjoista ja niiden muutoksista uudella ajalla*. Helsinki: WSOY.