Gerrit Bloothooft

✉ G.Bloothooft@uu.nl

ⓘ https://orcid.org/0000-0002-3124-9541

⌂ Utrecht Institute of Linguistics, Utrecht University,

🌐 Utrecht, The Netherlands

David Onland

✉ D.Onland@uu.nl

ⓘ https://orcid.org/0000-0001-8098-0608

⌂ Utrecht Institute of Linguistics, Utrecht University,

🌐 Utrecht, The Netherlands

Online Exploration of Given Name Popularity in the Netherlands Since 1790

Abstract

Given names are an expression of identity and reflect socio-cultural and linguistic relationships in a society. The variation in popularity of given names over time can be studied in the Netherlands from 1790 onwards, based on the digital civil registration (21.9 million births, 1880–2017) and digitized marriage certificates (resulting in 4.04 million births of brides and grooms, 1790–1880). Data is accessible online in the Dutch corpus of given names which provides information on 480,851 different names. Some characteristics of long-term popularity changes for individual names, but also the possibility to show the popularity of sets of names are demonstrated. For the latter, there is the option to use regular expressions to define a pattern that names should match. Examples are given to clarify the possibilities.

1. Introduction

Given names are a reflection of a society's socio-cultural and linguistic relationships. Since the registration of names, sex, family relations and places of birth and residence is the backbone of a nation's administration, and is usually archived very carefully, this information is potentially a very rich source for socio-onomastic studies – if it can be accessed. Examples are studies on commemorative naming of deceased relatives in England and Wales since 1837 (Bush, 2019), on modeling fashion in given names, based on data since 1880 from the USA and France (Berger & Le Mens, 2009), and from the USA, France and the Netherlands (Coulmont et al., 2016), and on the distribution of groups of given names in the Netherlands since 1880 (Bloothooft & Schraagen, 2014).

For the Netherlands, we have now created the unprecedented possibility to access the annual given name frequency online from 1790 to 2017 for 481,432 different given names, and their geographic distribution for the current population. The information can be searched by given name and by regular expressions which is a powerful tool to study popularity and geographic distribution for sets of given names that fulfill a pattern.

We will explain how data is derived from the digital civil registration (21.9 million births 1880–2017) and digitized marriage certificates (resulting in 4.04 million births 1790–1880). In both cases population reconstruction techniques were needed. The precision of both approaches was confirmed by the continuity of the data in 1880, where the historical and modern data meet.

Popularity features will be demonstrated with emphasis on the 19th century and their continuation until today, both for individual names and for sets of names that fulfill some pattern as defined by a regular expression.

2. Data sources

To establish given name popularity in the Netherlands after 1880, we used the digitized Dutch Civil Registration (Basisregistratie Personen, BRP) from which it is legally allowed to request data for scientific research. For onomastic studies we were allowed to obtain the given name and year of birth for all those who hold Dutch citizenship and were born in the Netherlands, as of January 1, 2018. Since the digitized Civil Registration was introduced in 1994, all 17.86 million people who were alive in or after that year are included. As the given names and year of birth of the parents of each person were available as well, this opened the possibility to reconstruct individuals who died before 1994. This required the deduplication of parents who were mentioned in records of several children. Another 4.04 million people could be added in this way for a total of 21.9 million individuals. In all, this is a substantial sample of the given names of the Dutch population after 1880. Figure 1 shows the annual number of given names (both male and female), with, as a reference, the number of births given by Dutch Central Bureau of Statistics (CBS, 2001, p. 14). Recent small differences between figures originate in our requirement of Dutch citizenship. The popularity of a given name in a given year can be computed as a separate percentage for males and females. For an estimation of the total number, this percentage is applied to the total number of births from CBS.

The Dutch vital registration, which began in 1811 and contains birth, marriage and death certificates, can be used as the source for given names of births before 1880. Since the 1990s, numerous volunteers have digitized key information from these certificates (to be found at https://www.wiewaswie.nl). Because the birth certificates have so far only been partly digitized, marriage certificates, for which the digitization is almost complete for the period 1811–1937, have been used. The given names of bride and groom and their age (and estimated year of birth) were used to obtain an annual sample of births with given names for the period 1790–1880. Note that most people born after 1790 marry after 1811 when they are older than 21 years. A concern with this approach is multiple marriages and multiple mentions as bride or groom. We therefore removed records with the same given name, surname, gender, and year and place of birth. However, for a limited number of cases of popular full names (*Jan de Jong*) in larger municipalities or incorrect estimation of year of birth this may have led to
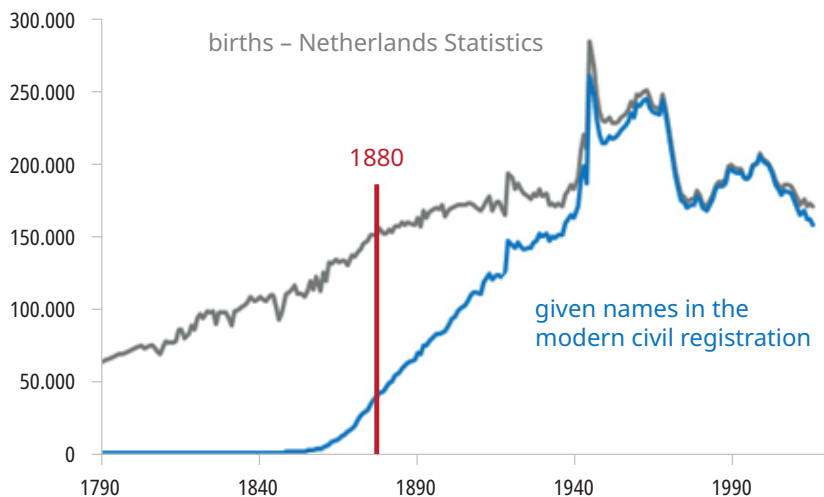
Figure 1. Annual number of births with known given name as derived from the modern Dutch Civil Registration, with the total number of births from CBS as a reference
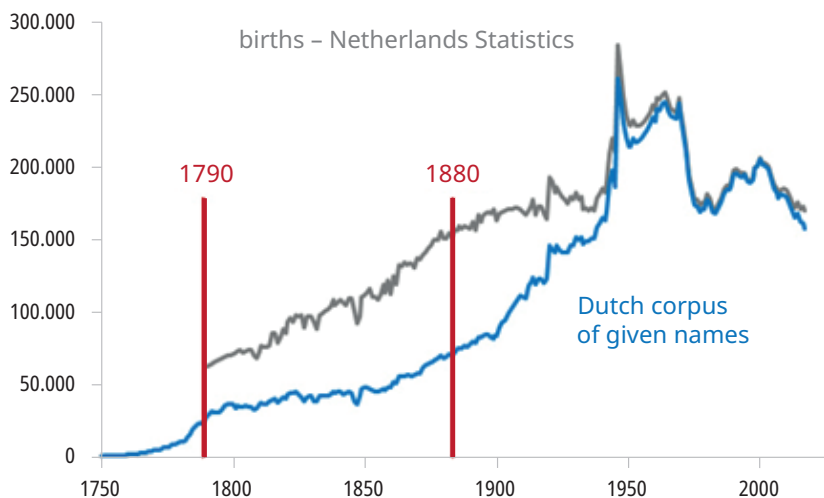
Source: own work.

Figure 2. Annual number of births in the Dutch corpus of given names (1790–2017), derived from marriage certificates (1790–1880) and the civil registration (1880–2017), with the total number of births from CBS as a reference

Source: own work.

incorrect deletion. Our requirement of Dutch citizenship in the civil registration is approximated in the marriage registers by birth in the Netherlands for the bride and groom. In total, for the period 1790–1880 the given names of 4.04 million persons were obtained. In combination with the data derived from the civil registration, Figure 2 shows the annual number of births with known given name for the full period 1790–2017. In the 19th century, the annual percentage of births with a known given name is over 40%, but it should be kept in mind that in the Netherlands child mortality (between 0–5 years) was between 22% and 35% in that period (Tabeau et al., 1994). This implies that the Dutch corpus of given names contains more than half of the population over five years old in the 19th century, up to almost the full population after 1945.

The information from the Dutch corpus of given names can be found online in both Dutch and English from https://www.meertens.knaw.nl/nvb/. It provides the popularity of 480,851 different given names as initial name and/or middle name, both as a percentage and (estimated) number. Only a limited number of 35,593 given names have a presence before 1880, out of which 4,134 do not appear after 1880. The major variation in given names arises over the course of the 20th century. In addition, annual popularity lists (top-100) are provided from 1880 onwards. For the current population the geographical distribution of a name is presented at municipality level. For 20,236 given names an explanation and background information are provided as originally compiled by Van der Schaar (1964). There is a search function with the options: *is equal to*, *begins with*, *ends in*, *contains*. There is also an advanced search by which popularity or a geographic spread of a set of names that fulfil a pattern as specified by a regular expression can be shown. We will demonstrate the possibilities of the latter in more detail in section 4. Although all given names are presented, to protect privacy, the details are only given when there are five or more name bearers at any presentation level.
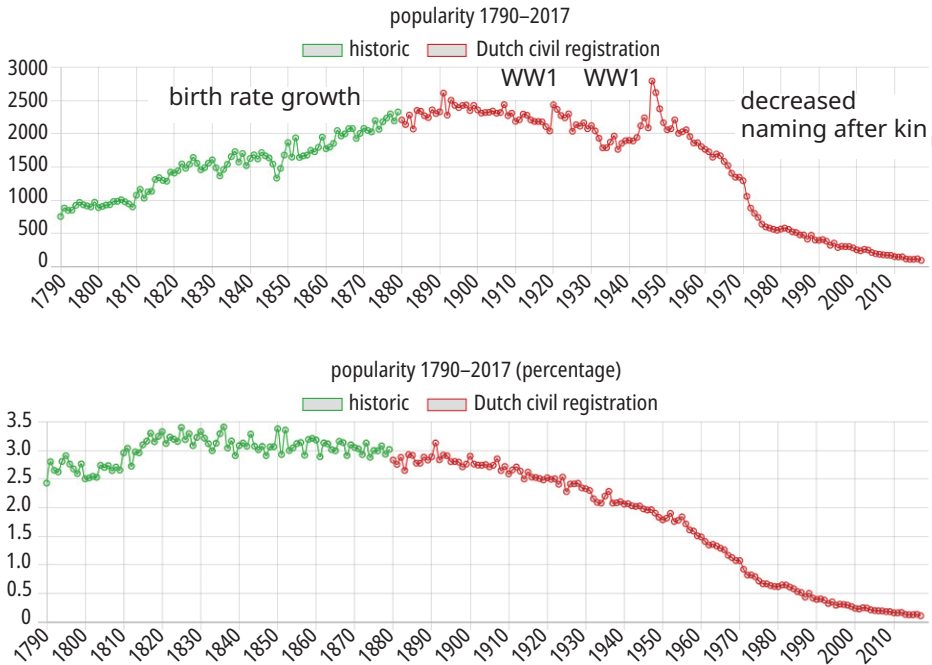
3. Some issues in given name popularity

Given name popularity can be given as annual number or percentage.[1] The former is dependent on variation in the birth rate as can be seen for the traditional male name *Gerrit* in Figure 3. It shows the effect of birth rate growth in the 19th century and post-war baby booms. The actual decline in the popularity of *Gerrit*, and traditional Dutch names in general, is much more gradual and is already underway at the end of the 19th century, being related to the decrease in naming after kin (mainly grandparents). Note that the continuity of historic and modern data in 1880 is very good.

For Catholics, the traditional Dutch name *Gerrit* (from Germanic *Gerhard*) is usually registered as the Latinized form *Gerardus* for which the relative popularity is shown in Figure 4. In 1811, the start of the civil registration, there is a sharp decline in popularity. From then on it is a civil servant who registers the name on the basis of the father's statement, while before it was the priest who wrote the name in the book of baptism. The decline suggests that the father preferred a non-Latinized form, as can been seen in the increase after 1811 for *Gerrit* in Figure 3. The person who registers a child can therefore have a considerable influence on the form of the name, inducing biased popularity. In contrast to *Gerrit*, *Gerardus* shows a gradual increase until the 1950s. After that date the fall is strong, due to secularisation and the abandonment of the practice of naming after grandparents. This pattern is found for many Latinized names. It should be noted that Latinized names, especially those ending in *-us*, are rarely in daily usage, but abbreviated to, for instance, *Gerard* in the case of *Gerardus*.

Sometimes, a name in decline can regain popularity. This is shown in Figure 5 for *Sara*, a name that became fashionable this century after a long decline until 1970. Another observation is the mismatch between historic data and modern data in 1880. This originates in the Holocaust during which a large part of the Jewish population was killed and is therefore not part of our reconstructed population sample. As *Sara* is the name of 2,682 Dutch females

---

[1] Rank is another presentation type, which can be accessed as annual top-100 lists at municipality, regional and country level (after 1880).

popularity 1790–2017

□ historic   □ Dutch civil registration

birth rate growth   WW1   WW1   decreased
                                                 naming after kin

popularity 1790–2017 (percentage)

□ historic   □ Dutch civil registration

Figure 3. Popularity of the male name *Gerrit* (1790–2017), both in numbers (top panel) and percentages (bottom panel). Green is data from marriage registers, red data is derived from the modern civil registration

Source: own work.

popularity 1790–2017 (percentage)   Gerardus

□ historic   □ Dutch civil registration
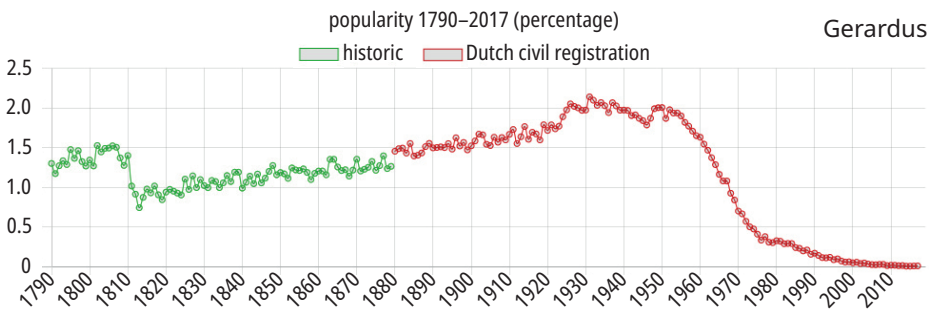
Figure 4. Relative popularity of the Latinized male name *Gerardus* (1790–2017)
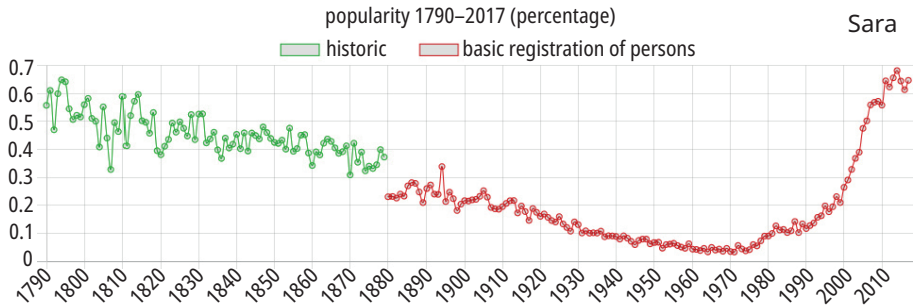
Source: own work.

who died in the Holocaust,[2] its relative presence is therefore underestimated. Since the historical marriage registers date from before the Second World War, the data derived from these is more representative.

The popularity of a middle name can be worth studying, as it may be used for specific purposes. But the interpretation of the popularity of a middle name requires attention, as is shown in Figure 6 for the male name *Hubertus.* The popularity as a first name demonstrates the pattern shown earlier for *Gerardus*: a steady increase from 1790 onwards and a strong decline after 1960. *Hubertus* as a middle name starts at the end of the 18th century. It reaches a peak at the end of the 19th century, with a steady decline afterwards. However, the increase in its popularity is not only the result of the attractiveness of the name itself, but of the popularity of giving more than one name to children (Bloothooft & Onland, 2016). Especially in the south of the province of Limburg, *Hubertus* is a very popular saint, and a second or third name opens the possibility to express this popularity (whereas as first given name it is traditionally used to name after grandparents). This also explains why *Hubertus* is much more popular as a middle than as an initial name. The drop in popularity after 1918 is a reaction to the proposal to raise tax for more than one name (which never came into effect), as discussed in Bloothooft and Onland (2016).

The popularity of a given name is rarely homogenously distributed over a country. Its geographic distribution may reveal origins of popularity, for instance related to religion, social differences, regional languages or dialects. The corpus of Dutch given names has the option to present the geographic distribution of (sets of) names based on the places of birth for the population in 2017. The presentation can be shown as a number or a percentage per municipality. An example is given in Figure 7 for the geographic distribution of the given names *Gerrit* and *Gerardus*. It shows the division of the Netherlands in the Catholic south and north-west (*Gerardus*), and the Protestant middle and north of the country (*Gerrit*).

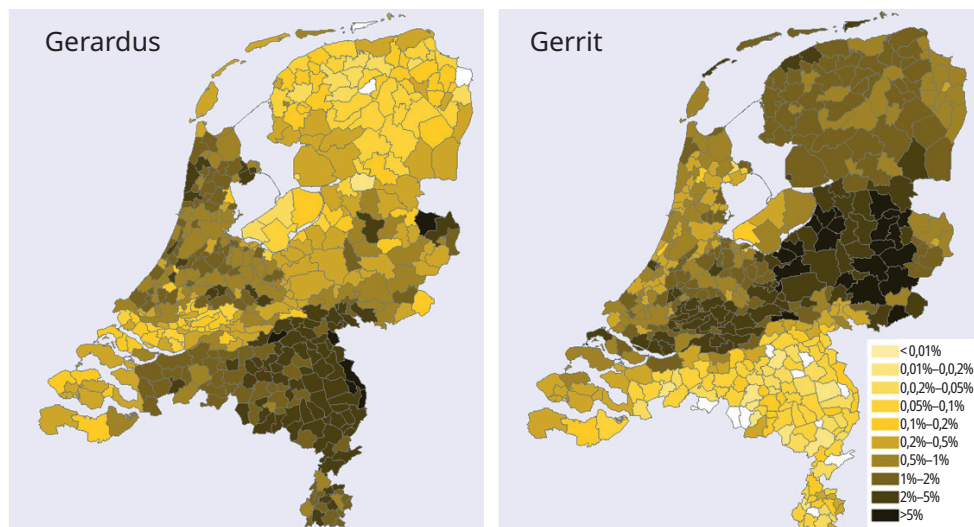[2] As registered at https://www.joodsmonument.nl/

Figure 5. Relative popularity of the female name *Sara* (1790–2017)

Source: own work.
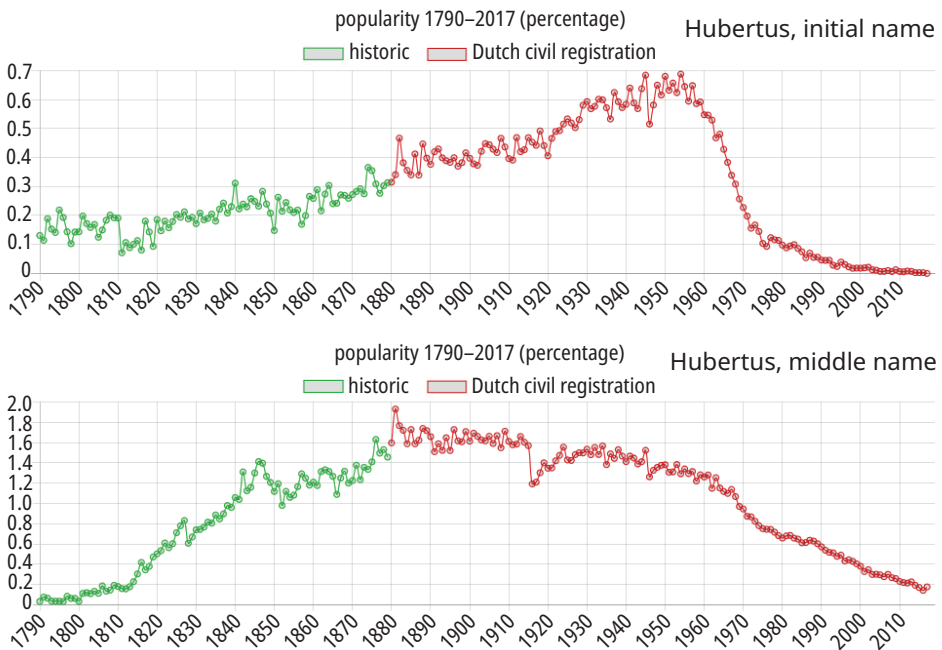
Figure 6. Relative popularity of the male name *Hubertus* (1790–2017), as an initial name (top panel) and as a middle name (bottom panel)

Source: own work.

Figure 7. Geographic distribution of the given names *Gerardus* and *Gerrit*, as a percentage per municipality, based on places of birth of the population in 2017
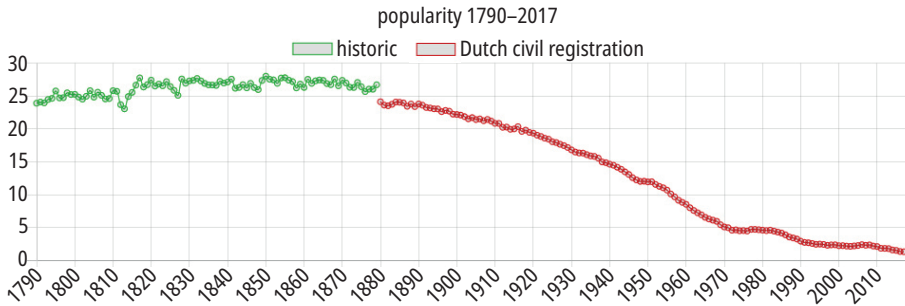
Source: own work.

4. Joint popularity of given names using regular expressions

Besides the study of the development of the popularity of a given name, it is often of interest to study the joint popularity of names that match a pattern, such as having the diminutive *-je*, or all variants of names that begin with *Wil*. For the description of such a pattern there is the powerful technique of regular expressions, widely used in search engines. The online corpus of Dutch given names has an advanced search option that uses regular expressions.[3] Here we have given some examples of the possibilities.

If we want to show the use of the diminutive *-je*, the expression je$ is sufficient, where $ equals the end of a name. This results in 7,908 names, listed by frequency (either male or female). The joint popularity for female names

---

[3] A short description of the usage of regular expressions, geared towards given names, is given at https://www.meertens.knaw.nl/nvb/regexp

on *-je* is shown in Figure 8. Whereas in the 19th century a quarter of women had a name with such a diminutive, it is currently less than one percent. Since the diminutive usually turns a male name into a female form (*Dirk > Dirkje*), the decline can be interpreted as an expression of female independence that was already underway in the late 19th century.

Figure 8. Joint relative popularity of all female given names ending in the diminutive *-je* (1790–2017)
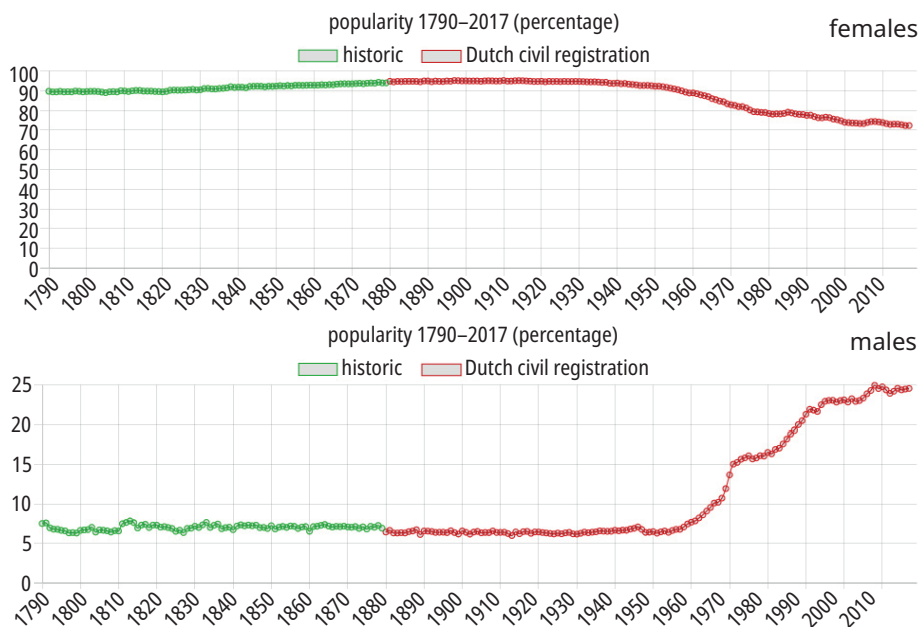
Source: own work.

Table 1. Examples of major formal regular expression descriptions, useful for describing patterns in names (case-insensitive)

^jan	starting with *jan*
ke$	ending with *ke*
(jan|willem)	containing *jan* or *willem*
[ptk]	containing one of the characters *p, t* or *k*
m*	containing any number of sequential characters *m*, including zero
m+	containing a sequence of at least one character *m*
r?	optional character *r*
.	any character

Table 1 lists the major formal regular expression options for describing patterns in names. All options may be combined. As an example, the expression ([aeiouy]\|ij)$ describes all names ending in a vowel. This results in 196,709

different female names and 84,501 different male names for which the joint popularity is shown in Figure 9. Between 1790 and 1960 over 90% of all women have a name that ends in a vowel, while for men this is only 7%, showing a clear gender difference. But over the last 50 years the percentage decreased to 70% for women while increasing to 25% for men, resulting in a diminished likelihood that a name ending in a vowel indicates a woman.


Figure 9. Joint popularity of 196,009 different female names (top panel) and 84,501 different male names (lower panel) ending in a vowel

Source: own work.

It is also possible to derive the joint popularity of a specific set of names. For instance, for the French male names *Jean*, *Pierre*, *Guillaume*, *Henri* and *Jacques* as ^(Jean|Pierre|Guillaume|Henri|Jacques)$ or the French female names *Marie*, *Jeanne*, *Jacqueline* and *Anne* as ^(Marie|Jeanne|Jacqueline|Anne)$ for which the popularity is given in Figure 10. *Jean*, *Pierre*, *Guillaume*, *Marie* and *Jeanne* were very popular in around 1790, but were in strong decline after the French revolution in 1789 and the escape of Willem V, Prince of Orange,

from the Netherlands to England in 1795. This was the start of the Batavian Republic, which was strongly influenced by the French empire, of which it even became part in 1810. In 1811, civil registration was started on the French model with French certificate texts, presumably leading to an increase in the registration of French versions of Dutch names (*Jan > Jean*, *Pieter > Pierre*, *Willem > Guillaume*, *Maria > Marie*, *Jannetje > Jeanne*). After Napoleon's final defeat in 1815, French names became entirely unfashionable. During the 19th century the French male names *Henri* and *Jacques* gradually became more fashionable but declined after 1970. In contrast, French female names *Jacqueline* and *Anne* are examples of time-specific fashion names after 1950.
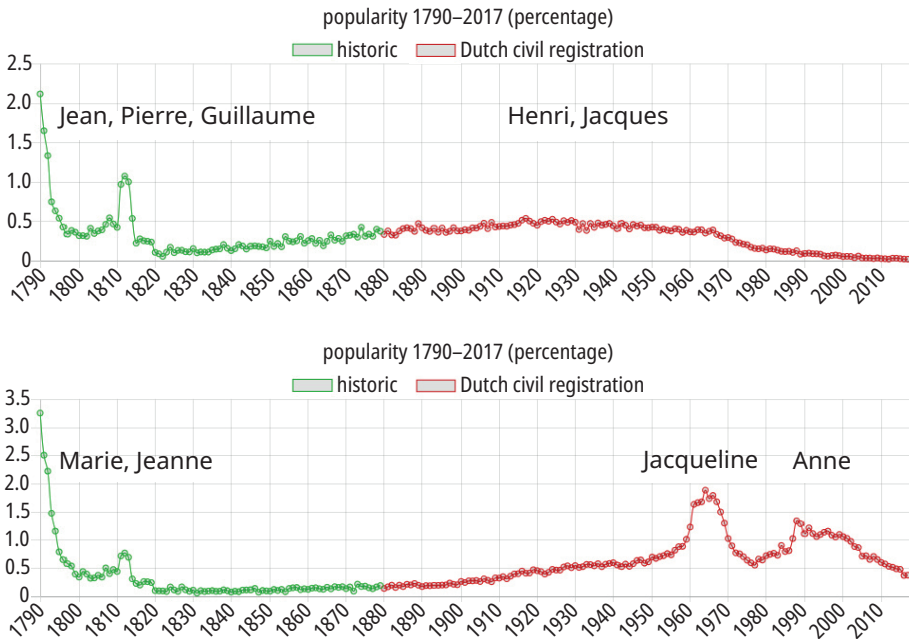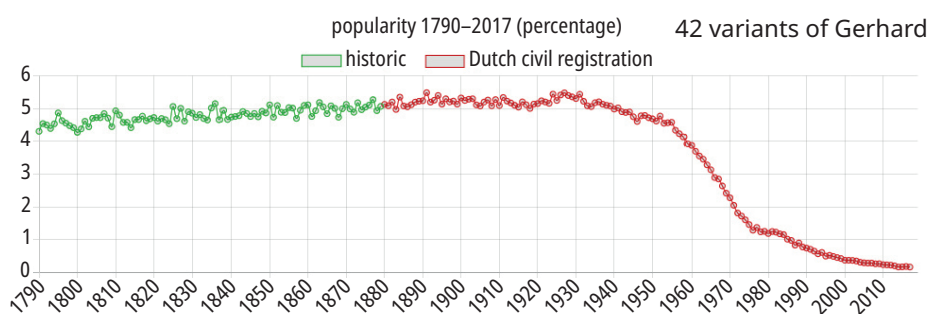

Figure 10. Joint relative popularity of the French male names *Jean*, *Pierre*, *Guillaume*, *Henri* and *Jacques*, and the French female names *Marie*, *Jeanne*, *Jacqueline* and *Anne*

Source: own work.

More complex regular expressions can be used to describe variants of a given name. For instance, ^Ger*h?[ia]r?[td]u?s?$ summarizes 42 variants of the

Germanic name *Gerhard*. Most popular are *Gerardus, Gerrit, Gerard, Gerhardus*, and *Gerhard*, but rare variants such as *Geradus, Gerit, Gerrard, Geart, Gerad*, etc. are also included in the expression. The joint popularity (Figure 11) shows stability until 1940 and the following decline afterwards – as with most traditional names. It is of interest that the major contributions from *Gerrit* (Figure 3) and *Gerardus* (Figure 4) show variation in popularity over time, while the joint contribution of all variants of *Gerhard* is much more stable. We encourage the reader to try the possibilities of regular expressions on https://www.meertens.knaw.nl/nvb.
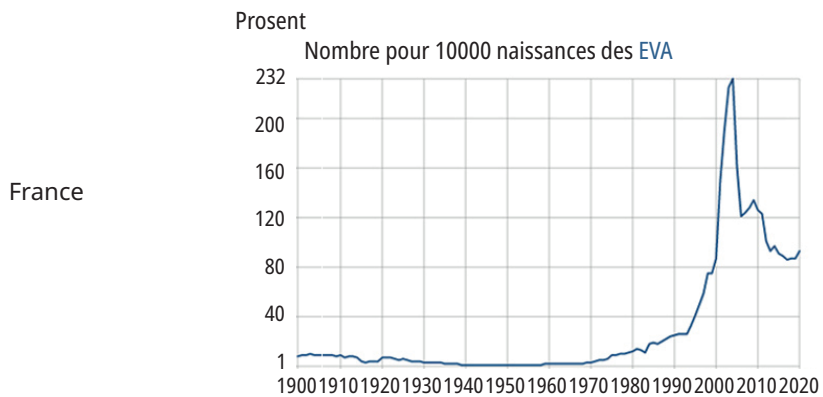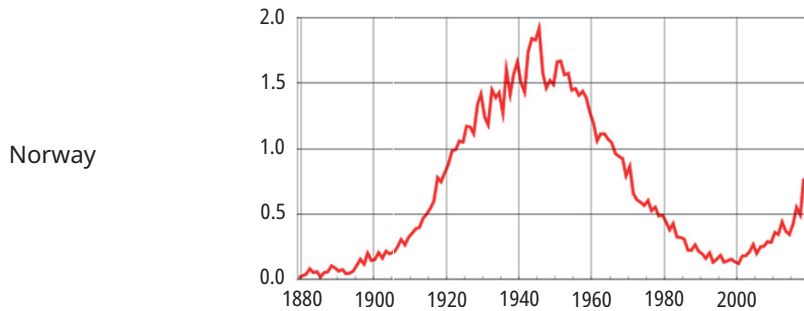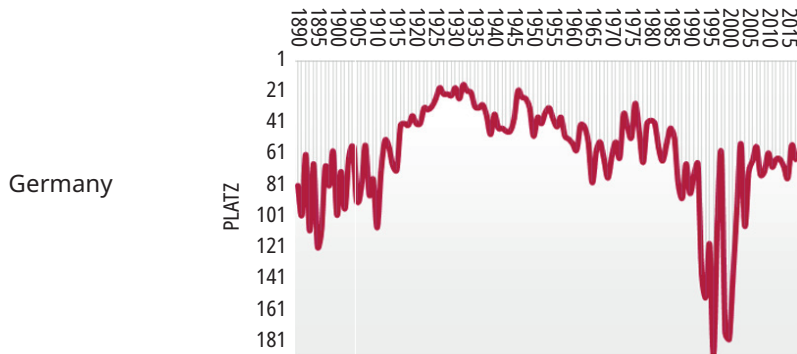

Figure 11. Joint relative popularity of 42 variants of the Germanic name *Gerhard*

Source: own work.

5. Popularity of given names elsewhere

For various countries, online data is available on given name popularity, but over various time ranges and often only for top names and with varying ways of presentation. Data documentation is very limited in most cases.[4] It is of interest to compare the development of the popularity of a given name in different countries. This is done in Figure 12 for the popularity of the female

---

[4] The websites https://nameberry.com/popular-names and https://www.babynamewizard.com present entries to given name popularity in many countries.

Babies named 'EVA' per million babies.   per million births

USA

Eva

	4,000
	3,500
	3,000
	2,500
	2,000
	1,500
	1,000
	500

1880s 1890s 1900s 1910s 1920s 1930s 1940s 1950s 1960s 1970s 1980s 1990s 2000s   2019

Germany

PLATZ

1890 1895 1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015

1
21
41
61
81
101
121
141
161
181

Norway

2.0
1.5
1.0
0.5
0.0

1880   1900   1920   1940   1960   1980   2000

Prosent

Nombre pour 10000 naissances des EVA

France

232
200
160
120
80
40
1

1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020

Figure 12. Relative popularity of the female name *Eva* in the USA, Germany (rank), Norway, France and the Netherlands, on a lined-up time scale

Sources: USA – https://www.ssa.gov/oact/babynames/, graph taken from https://www.babynamewizard.com; Germany: https://www.beliebte-vornamen.de; Norway: https://www.ssb.no/en/befolkning/navn/statistikk/navn; France: https://www.insee.fr/fr/statistiques/3532172; The Netherlands: https://www.meertens.knaw.nl/nvb

name *Eva* in the USA, Germany, Norway, France and the Netherlands. There are remarkable differences. In the USA *Eva* was very much in fashion in the beginning of the 20th century, while in Germany the highest rank was around 1930, followed by Norway where the peak came somewhat later in 1940. In France and the Netherlands, *Eva* had little presence until 1980, but peaked in France in 2004 (just a few years after the name reached its lowest popularity in Germany). In the Netherlands there was a steep rise to the high popularity that continues to today. We do not aim to explain these differences, but the observations could stimulate research into the mechanisms behind the international spread of the popularity of given names.

6. Epilogue

The possibility to acquire onomastic ground data from the civil registration is rare. When in addition, as in the Netherlands, there is a long-term effort to digitize historical vital certificates by volunteers, this creates an unprecedented opportunity for studies of given names over centuries with high precision. Taking

all possible precautions in relation to privacy, this data can be made available to the public. At the very least this is of interest to prospective parents search-ing for names for their children and wanting to know all given names that are already formally accepted in the Netherlands and their popularity. No doubt this contributes heavily to the over 10 million page views per year of the corpus of Dutch given names, along with the civil servants responsible for child registra-tion who regularly check the corpus. Besides the presentation of the given names of a very large sample of residents with Dutch nationality, the online corpus pre-sents the possibility for onomastic studies, where especially the option to select sets of names using regular expressions is very valuable. It is also important to take a broader perspective in the understanding of the world-wide spread of popular names. To that end we advocate international co-operation.

References

Berger, J., & Le Mens, G. (2009). How adoption speed affects the abandonment of cultural tastes. *PNAS*, *106*(20), 8146–8150. https://doi.org/10.1073/pnas.0812647106

Bloothooft, G., & Onland, D. (2016). Multiple first names in the Netherlands (1760–2014). *Names*, *64*(1), 3–18. https://doi.org/10.1080/00277738.2016.1118860

Bloothooft, G., & Schraagen, M. (2014). Name fashion dynamics and social class. In J. Tort i Donada & M. Montagut i Montagut (Eds.), *Names in Daily Life. Proceedings of the XXIV International Congress of Onomastic Sciences* (pp. 419–426). Barcelona: Generalitat de Catalunya. https://doi.org/10.2436/15.8040.01.47

Bush, S.J. (2019). Re-using the names of newborns: Symbolic reincarnation in an age of infant mortality. *Names*, *67*(2), 100–112. https://doi.org/10.1080/00277738.2018.1536186

CBS – Centraal Bureau voor de Statistiek (2001). *Tweehonderd jaar statistiek in tijdreeksen 1800–1999.* Voorburg/Heerlen: Uitgeverij Stichting Beheer IISG. https://www.cbs.nl/nl-nl/publicatie/2000/51/200-jaar-statistiek-in-tijdreeksen

Coulmont, B., Supervie, V., & Breban, R. (2016). The diffusion dynamics of choice: From durable goods markets to fashion first names. *Complexity*, *21*(S1), 362–369. https://doi.org/10.1002/cplx.21748

Tabeau, E., Willekens, F. & van Poppel, F. (1994). *Mortality in the Netherlands: The Database* (NIDI report 36). NIDI: The Hague.

van der Schaar, J. (1964). *Woordenboek van voornamen*. Aula: Utrecht.