

Personal names in The Netherlands: data sharing for international comparisons



Gerrit Bloothoof
Utrecht University, the Netherlands

contact:
g.bloothoof@uu.nl
www.gerritbloothoof.nl (in Dutch, with publications)

1

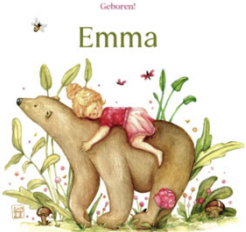
focus

- research questions that exceed borders
 - with an example
- personal names corpora
 - The Netherlands
 - international
- issues on data sharing

2

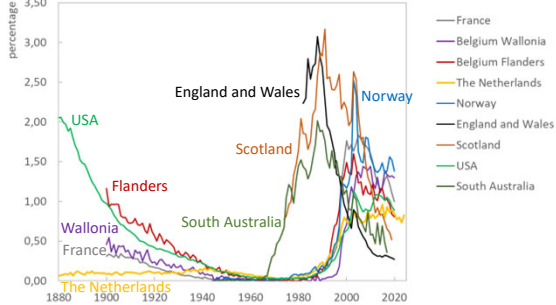
to set the stage

an international comparison of the popularity of the given name *Emma* since 1880



3

popularity of *Emma* in various countries



4

some research issues that exceed borders

- diffusion of fashionable given names
 - modelling / power law properties / hapax onomena
- geographic spread of names
 - role of state (regional) borders and language borders in name choices (cultural borders?)
- role of social status and religion in name choices
- migration onomastics

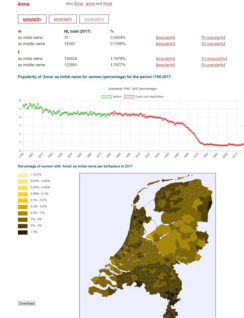
all issues: over time

5

Dutch data on contemporary given names (1880+)

acquired from the civil registration *on a legal basis* for onomastic research (2006, 2011, 2017, 2022)
(full population, 444,000 names, 24.9 million persons)

- internal id
- ids of parents
- gender
- all given names
- date, place and country of birth
- date of death or migration
- postal code (municipality of living)
-



online (data per given name): nvb.meertens.knaw.nl

privacy issues in publication (The Netherlands)

data presentation

- no individual may be identified
- an individual may not be traceable

ownership

- one cannot own a given name:
 - there is no intellectual property of a given name
- transparency (equality for all):
 - all given names and surnames should be known to everyone

13

data sharing

- no individual data
- aggregated data
 - on contract basis

14

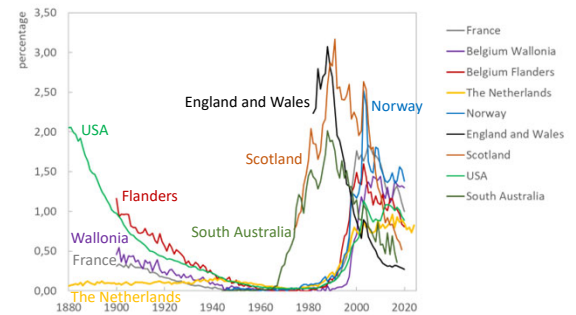
example: modeling given name popularity (for Emma in USA, France, the Netherlands, Scotland)

model, based on an analysis of the popularity of 20,000 new given names introduced in The Netherlands between 1920-1960 (with data until 2017)

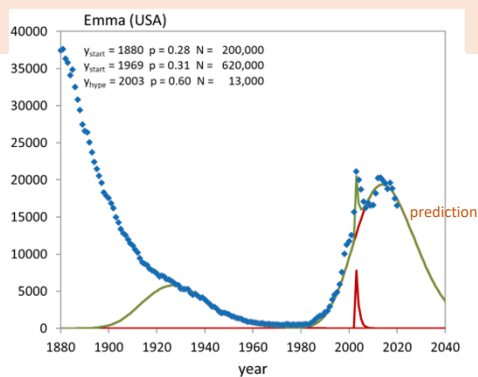
- **word of mouth process (slow wave)**
 - starting year
 - parents' copy probability (based on intrinsic name features)
 - number of transitions through social networks (total number of namings)
- **hype (fast peak)**
 - starting year
 - parents' copy probability (high)
 - behaves as a single network

15

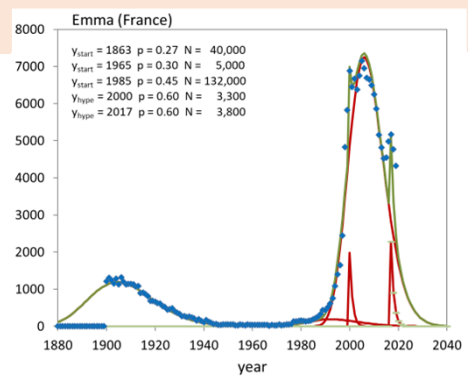
popularity of Emma in various countries



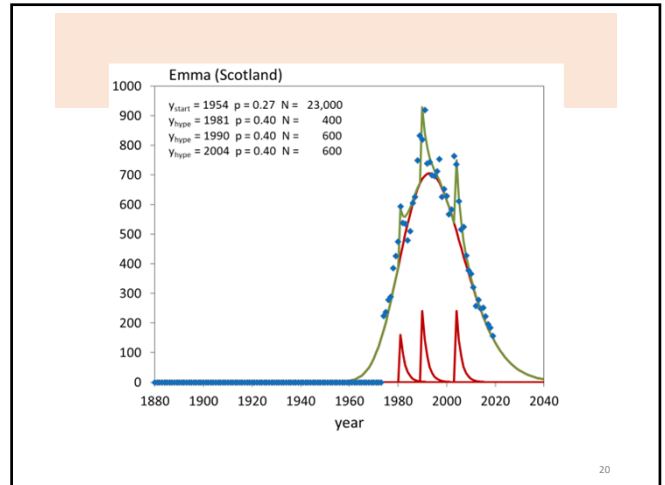
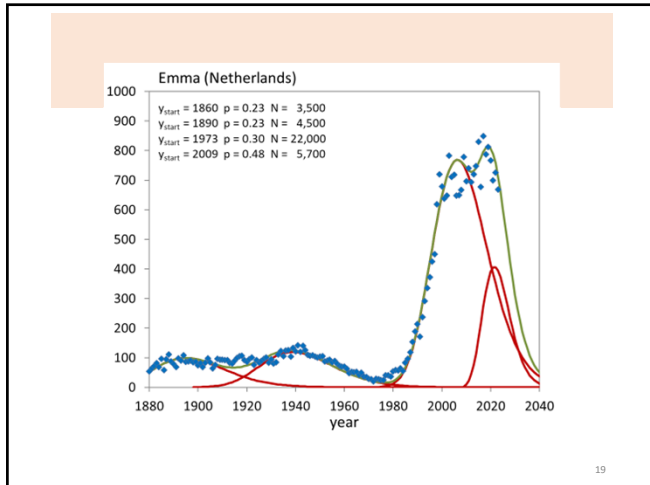
16



17



18



pros, cons and questions in popularity modelling

strong:

- the model is based on precise dates (days!) of name giving
- the model unravels very well underlying waves and hypes
- the model is widely applicable (for many countries)

weak:

- manual model optimization

pending questions:

- *the societal origins* of popularity waves and hypes
- *why* given names differ in popularity
- *why* some names become internationally popular

21

general conclusions

- (Western) personal name choices may or may not cross borders: its study requires **international efforts and cooperation**
- there is a need for a **repository** with available personal name data per country (over time): how do we organize this (ICOS/ANS)?
- we could explore the possibilities for **data sharing** (at aggregated level)

22

abstract

In The Netherlands, legislation makes it possible to get access to personal name data from the civil registration for onomastic research in selected institutions. On this basis, key name data for the full population have been made available on line in the corpora of given names (nrv.meertens.knaw.nl) and family names (www.cbgfamilienamen.nl/nfb/). Whereas a part of name choices may be typically Dutch, in many cases these choices result from diffusion processes with international impact. To understand these processes of human communication and societal influences, international contrastive comparison of name frequency and geographic spread (both historically and modern) is vital. Sharing name data, however, is not trivial because of legal limitations on access and distribution related to privacy. Nevertheless, a first step forward could be made by developing an inventory of person name data that are currently made available in countries through national bureaus of statistics or other official bodies. Whereas person name data for the Netherlands cannot be shared at the individual level, this is possible for aggregated data on a contract basis. If comparable options exist in other countries this would open the way for joint, international research on onomastic processes that exceed borders. Examples will be presented for given name popularity in The Netherlands, USA (from 1880 onwards), France, Belgium (Wallonia and Flandres) (from 1900), Norway (from 1945), South Australia (from 1950), Scotland (from 1974), England and Wales (from 1982) – and their modelling by word of mouth processes and hypes. Other research option examples are the power-law behaviour of name choices and the role of hapax onyms, given names that are only given once (in a year).

23