Who’s who in the Netherlands
Gerrit Bloothooft

Onomastic topics

Identity in names
Names for identification

Identity in names

Names of children:
Sociolinguistics
Sociology

Parents and children

Parents prefer:
- Same type of name
- Same number of names

From tradition to fashion

Big data from civil registration:
23 million individuals

Popular in the same family

Lars - Niels
Martijn - Jeroen
Bas - Tom
Maarten - Wouter
Martijn - Sander
Bas - Tim
Mike - Roy
Daan - Koen
Mike - Nick
David - Ruben

Bram - Daan
Martijn - Jasper
Mark - Linda
Johannes - Maria
Roy - Kim
Bart - Koen
Bart - Tom
Jeffrey - Wesley
Patrick - Chantal
Johanna – Maria
name groups

Lars - Niels
Martijn - Jeroen
Jeroen
Maarten - Wouter
Martijn - Sander
Bas - Tim
Mike - Roy
Mike - Nick
David - Ruben
Bram - Joran
Martijn - Jasper
Mark - Linda
Johannes - Maria
Roy - Kim
Bart - Roen
Bart - Tim
Jeffrey - Wesley
Patrick - Chantal
Johanna - Maria

mapping of name groups

name groups and social stratification

Wegener Direct Marketing
- 1 million households
- consumer segmentation
- family composition

number of names

Hendrik
Jan Willem
Bart Johannes Maria
Willem-Alexander Claus George Ferdinand

parental preferences over time?

Grietje
Johanna Maria
Anouk Johanna Maria
Alexia Juliana Marcela Laurentien

number of names 1750-2014

1916 first name tax proposal
names for identification

computational linguistics
artificial intelligence
social history
demography

variants and errors

Willem
Wilhelmus
Guillaume
Willenis
W8llem
Gilliaumme
Qwillem

automatic learning of variation

person records with:
• first name
• approximate year of birth
• first name father
• surname father
• first name mother
• surname mother

assumption: all information except one name is sufficient for identification

collect:
variation in unused name
Zipf’s law applies

189,000 singular first names
  44,000 standardized (coverage 98.8%)
  2,000 standards anticipated

681,000 surnames
  126,500 standardized (coverage 90.6%)
  20,000 standards anticipated

In conclusion

The study of proper names in big data is vital to demonstrate aspects of social stratification to identify people in the past.