

In de naam van ...

8. De voornamenkaart van Utrecht

Gerrit Bloothoof

Tussen 1999 en 2001 heb ik zeven artikelen geschreven voor Link over naamkundige onderwerpen. Daarna werd het stil. Maar dat was schijn, want er is sindsdien met Emma van Niferick en Doreen Gerritzen hard gewerkt aan het boek 'Over voornamen', dat dit voorjaar bij Het Spectrum verschijnt. We hadden via de Sociale Verzekeringsbank de beschikking gekregen over vrijwel alle voornamen die sinds 1983 aan 4.2 miljoen kinderen zijn gegeven. Dat zijn zo'n 177.000 verschillende namen, waar geweldig veel over te vertellen en aan te onderzoeken is. Maar veel daarover staat in het boek en dat zal ik hier niet herhalen.

Het hoogtepunt van het boek is voor mij de presentatie van de voornamenkaart van Nederland. Daar liggen analytische procedures aan ten grondslag die erg interessant zijn, maar zich niet in detail laten presenteren in een publieksboek. Daarom doe ik dat hier. De ontwikkeling van de kaart begint eigenlijk op een bijeenkomst bij de Sociale Verzekeringsbank (SVB) waar Doreen Gerritzen en ik een presentatie hielden over de analyses die we hadden gemaakt van voornaamgegevens die we tot dan toe van de SVB hadden gekregen namelijk over 1998 en 1999. We hoopten hierdoor goodwill te kweken bij de SVB opdat we alle voornamen vanaf 1983 zouden kunnen krijgen. De mensen van de SVB waren erg enthousiast, en vertelden en passant dat ze naast voornamen ook informatie hadden over het gezin van het betrokken kind. Op dat moment onderkende ik onmiddellijk het belang van die informatie, omdat daarmee de samenhang van voornamen kan worden onderzocht. Tot onze vreugde kregen we na die presentatie niet alleen alle voornamen vanaf 1983, maar ook per kind de viercijferige postcode ten tijde van de geboorte en het gezinsnummer. Zo had ik de gegevens in handen, maar nu moest daaruit nog een voornamenkaart gedistilleerd worden.

Naamgroepen

Een probleem bij voornamen is, dat er in een betrekkelijk klein gebied zoals dat van de viercijferige postcode met gemiddeld zo'n 3 à 4000 inwoners, jaarlijks gemiddeld maar 60 kinderen worden geboren. Omdat ouders uit duizenden namen kunnen kiezen, zie je niet zo heel veel aan de verdeling van individuele voornamen per postcodegebied. En zelfs als je over een periode van 20 jaar alle namen samenneemt, is de frequentie per voornaam niet hoog. Eigenlijk zou je bepaalde voornamen willen samennemen onder de hypothese dat zulke voornamen zich op eenzelfde manier gedragen. Bijvoorbeeld omdat een bepaalde groep ouders een voorkeur heeft om bij de geboorte van een kind te kiezen uit een specifieke groep namen. Marokkaanse ouders geven hun kinderen vooral Islamitische, Arabische namen, en zullen niet kiezen voor Dirk of Klaas. Omgekeerd zullen gereformeerde ouders

op de Veluwe niet voor Mohamed kiezen, maar ook niet zo snel voor Kevin of Diederick. Om achter dit soort voorkeuren te komen is het belangrijk om te weten welke voornamen in eenzelfde gezin gegeven worden. Als een combinatie van twee namen vaak in een gezin voorkomt, dan is er een samenhang tussen die namen waarvan we mogen vermoeden dat die zijn oorsprong vindt in de sociale (opleiding, inkomen, geloof) en etnische achtergrond van de ouders. Kijk maar eens naar de 10 meest voorkomende namen van de broertjes en zusjes van Mohamed, Petrus, Jeroen en Kevin

	Mohamed		Petrus		Jeroen		Kevin	
rang	zusjes	broertjes	zusjes	broertjes	zusjes	broertjes	zusjes	broertjes
1	Fatima	Mohamed	Maria	Johannes	Linda	Martijn	Kelly	Dennis
2	Samira	Ahmed	Johanna	Martinus	Laura	Sander	Kim	Mike
3	Karima	Youssef	Anna	Cornelis	Suzanne	Dennis	Joyce	Jeffrey
4	Khadija	Rachid	Elisabeth	Wilhelmus	Sanne	Mark	Chantal	Michael
5	Nadia	Khalid	Catharina	Antonius	Marloes	Wouter	Melissa	Patrick
6	Hanane	Bilal	Cornelia	Franciscus	Marjolein	Bart	Mandy	Robin
7	Siham	Ali	Wilhelmina	Adrianus	Esther	Stefan	Denise	Wesley
8	Asma	Hamza	Petronella	Jacobus	Saskia	Erik	Melanie	Roy
9	Youssra	Yassine	Adriana	Paulus	Marieke	Jasper	Sharon	Danny
10	Ikram	Omar	Antonia	Gerardus	Anne	Maarten	Wendy	Nick

Hoe kun je nu op basis van deze gegevens namen groeperen? Eerst moet je een maat vinden voor de sterkte van de relatie tussen twee namen. Je kunt per voornaam (bv Jeroen) tellen hoeveel broertjes daarbij gevonden zijn, en hoeveel daarvan een bepaald naam (bv Martijn) dragen. Omgekeerd kun je tellen hoeveel broertjes Martijn heeft en hoeveel daarvan dan Jeroen heten. Ik koos als maat voor de sterkte van de relatie tussen de namen het geometrisch gemiddelde van beide fracties. Die maat voldoet aan de eis dat de sterkte symmetrisch is, dwz een gelijke waarde oplevert ongeacht of je van Jeroen of Martijn uitgaat. Loek de Groot van het SISWO (Instituut voor maatschappijwetenschappen) in Amsterdam, stelde later een maat voor die gebaseerd is op de berekening van het aantal broertjes van Jeroen die Martijn zouden heten als de ouders geen speciale voorkeur zouden hebben. De verhouding met het werkelijk gevonden aantal broertjes geeft dan de sterkte van de relatie aan, en is ook symmetrisch. In de praktijk voldeden beide maten even goed.

De sterktemaat valt alleen goed uit te rekenen wanneer de naamdragers een voldoende aantal broertjes en zusjes hebben. In ons geval voldeden 4000 voornamen daaraan. Elk van die voornamen staat in principe in een bepaalde relatie tot elk van de andere namen. Voor een aantal namen heel sterk en met andere namen nauwelijks. Nu is het zaak om de 4000 voornamen zo te groeperen, dat alle voornamen met krachtige onderlinge relaties bij elkaar komen in dezelfde groep. Nu zijn daarvoor wel standaard clusteringtechnieken beschikbaar, maar die vinden zulke grote aantallen niet zo fijn. Ik ontwikkelde daarom zelf een procedure waarin eerst de namen met een krachtige relatie in een groep werden gezet, waarna voor andere naamparen steeds werd bekeken bij welke groep de namen het

beste pasten, of dat het naampaar een nieuwe groep moest gaan vormen. In een aantal automatische correctierondes convergeerde de oplossing. Er werden weliswaar honderden groepjes gevonden, maar het leeuwendeel van de voornamen bleek met 25 groepen heel goed te beschrijven, zie Bloothoofd, G. (2001). Dat is een enorme datareductie. Om tot een voornamenkaart te komen was dat aantal echter nog aan de hoge kant. Daarom werden groepen die toch nog heel verwant waren, bij elkaar genomen tot de onderstaande acht groepen, plus een restgroep. De percentages geven het aantal kinderen dat tussen 1983 en 2001 een naam uit de groep kreeg. En klein kwart van de kinderen zit in een groep overige namen, omdat hun naam of tot de 173.000 laagfrequente, niet in de beschouwing betrokken namen, of tot de laagfrequente naamgroepen behoorde.

1. Traditionele namen in Latijnse vorm (Maria, Johannes, e.d.) [8%]
2. Traditionele namen in Nederlandse vorm (Hendrika, Jan, e.d.) [5%]
3. Friese namen (Fenna, Jelle, e.d.) [3%]
4. Voor-moderne Nederlandse en internationale namen (Suzanne, Jeroen, Linda, Mark, e.d.) [12%]
5. Buitenlandse namen (voor-modern, vooral Engels/Amerikaans) (Laura, Dennis, Melissa, Michael, Kim, Kevin, e.d.) [24%]
6. Korte moderne namen (Sanne, Tim, Anne, Bart, Britt, Emma, Anouk, e.d.) [13%]
7. Namen uit het Oude Testament, geschiedenis en cultuur, en natuurnamen (Esther, Daniël, Willemijn, Laurens, Iris, e.d.) [7%]
8. Arabische en Turkse namen (Nadia, Mohamed, Fatma, Mustafa, e.d.) [5%]
9. Overige namen [23%]

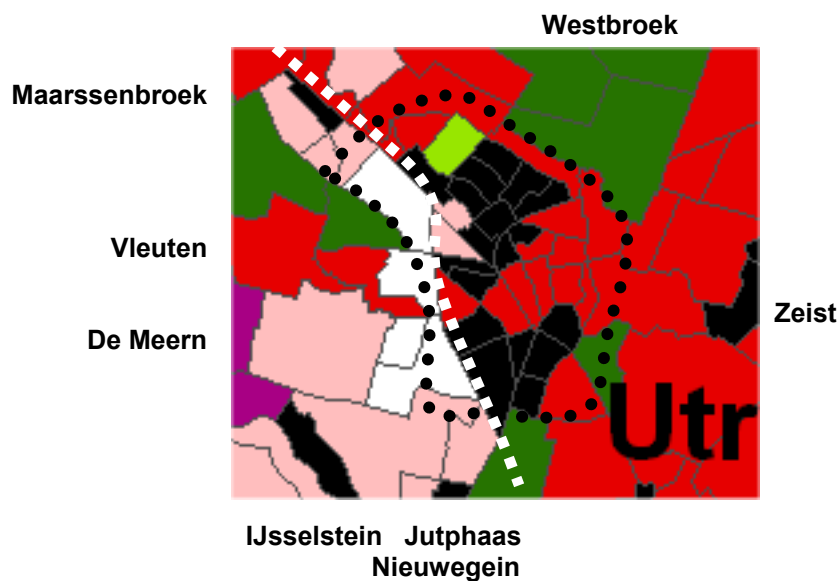
Het is bijzonder en uniek dat deze naamgroepen voortkomen uit de voorkeuren van ouders, en niet zijn samengesteld op grond van naamkundige overwegingen. Dat is nog voor geen ander land zo gedaan omdat de gegevens daarvoor simpelweg niet beschikbaar zijn. Ik suggereerde eerder een sociale en/of etnische achtergrond van deze naamgroepen. Voordat ik daarop inga moet eerst nog de geografische factor in kaart gebracht worden.

Voornamenkaart

Per postcodegebied kan voor elk van de acht groepen geteld worden hoeveel kinderen daaruit hun naam hebben gekregen. De bijbehorende acht percentages zijn veel bruikbaar dan de erg lage en onnauwkeurige percentages voor elke naam afzonderlijk. De verdeling van de percentages over de groepen – die we het voornamenprofiel noemen - kan erg verschillen tussen verschillende postcodegebieden. Als laatste stap moet daarom geprobeerd worden om een beperkt aantal representatieve voornamenprofielen te vinden, zodat elk postcodegebied aan precies één modelprofiel wordt gekoppeld. Pas dan kan de voornamenkaart getekend worden. Hiervoor biedt een bestaande clustertechniek uitkomst (k-means). Na wat experimenteren koos ik ervoor om 10 modelprofielen te onderscheiden. De clustertechniek koos vervolgens zelf de 10 modelprofielen, en wel zodanig dat de profielen van alle 3500 postcodegebieden de geringste afwijking hebben tot het meest bijpassende model. De modelprofielen hebben als kenmerk dat steeds één van de bovengenoemde groepen veel sterker dan gemiddeld aanwezig is. Daarnaast waren er nog twee overgangsprofielen.

Door elk modelprofiel een eigen kleur te geven, en daarmee de postcodegebieden in te kleuren, kan de voornamenkaart van Nederland getekend worden. Helaas wordt LINK niet in kleur gedrukt, dus die kunnen we hier niet laten zien. De kaart staat al wel in een artikel in *Onze Taal* (Bloothoofd et al, 2004). Hier lichten we alleen de stad Utrecht en omgeving eruit. Dat beperkt het aantal modelprofielen tot vier, en de kaart kan in zwart-wit worden weergegeven.

De stad Utrecht kent alleen maar twee extreme profielen: (1) het elitenamenprofiel met relatief veel namen uit het oude testament, geschiedenis en cultuur, en met natuurnamen, en (2) het autochtone-namenprofiel met veel Arabische en Turkse namen. In de voorsteden Maarssebroek en Nieuwegein vinden we het buitenlandse-namenprofiel, terwijl in de veenweidegebieden de traditionele namen nog overheersen. Probeer de onderstaande kaart eens te toetsen aan de eigen intuïties.



Voornamenkaart van de stad Utrecht en omstreken

Witte onderbroken lijn: Amsterdam-Rijnkanaal
 Zwarte stippellijn: Rondweg
 Witte gebieden zijn bedrijfsterreinen.

<i>Allochtone-namenprofiel</i> (zwart):	Overvecht, Zuilen, Ondiep, Pijlsweerd, Majellapark, Lombok, Transwijk, Kanaleneiland, Rivierenwijk, Tolsteeg, Hoograven, Sterrenwijk, en bepaalde wijken in Maarssebroek, Zeist, IJsselstein en Juthaas.
<i>Traditionele-namenprofiel met vooral Latijnse naamvormen</i> (groen):	Landelijke veenweidegebieden rond Westbroek en ten oosten en zuiden van de stad.
<i>Elitenamenprofiel</i> (rood):	Oog en Al, Centrum, Wittevrouwen, Tuindorp, Voordorp, Oudwijk, Schildersbuurt, Rijnsweerd, doorlopend richting Bilthoven en Zeist; Vleuten; Maarsse, Maarsseveen doorlopend langs de Vecht.
<i>Buitenlandsenamenprofiel</i> (roze):	De Meern, Maarssebroek, Nieuwegein.

De sociale en/of etnische achtergrond van de naamgroepen is onderzocht door het geloof, de urbanisatiegraad en het gemiddelde inkomen per postcodegebied (CBS-gegevens) in beschouwing te nemen. De traditionele namen worden vooral gegeven in de 'bible belt' die loopt vanaf Zeeland, de rivierenpolders, de Veluwe naar Noordwest Overijssel. Tot de Veluwe zijn de Latijnse naamvormen meer in zwang, noordelijker de Nederlandse. De kinderen zullen meestal overigens wel modernere roepnamen hebben. De elitenamen correleren hoog met het gemiddelde inkomen en we vinden ze vooral daar waar ook de gemiddelde woningprijs het hoogst is: de kuststreek, de Utrechtse heuvelrug, de Veluwezoom, en de rijke wijken van steden. De buitenlandse namen komen ofwel in wijken met het autochtone-namenprofiel voor en worden dan meestal door Surinamers en Antillianen gegeven, ofwel in de steden en het platteland van Holland, Zeeuws-Vlaanderen, Zuid-Limburg en de grensstreken van Groningen en Drenthe, en ze zijn dan waarschijnlijk populair bij de lagere inkomensklassen. Het allochtone-namenprofiel kent tot 50% Arabische en Turkse namen, veelal geconcentreerd in bepaalde wijken van steden. Profielen die niet voorkomen rond Utrecht zijn het korte-namenprofiel, populair in katholiek Nederland, het traditionele-namenprofiel met Nederlandse namen waarvoor we iets oostelijker op de Veluwe moeten zijn, het voor-moderne namenprofiel dat in Noordoost Nederland domineert, en ten slotte het Friese-namenprofiel waarvoor we uiteraard naar Friesland moeten.

Nabeschouwing

Het kernpunt van dit onderzoek was het vinden van het juiste aggregatieniveau van de gegevens langs de dimensies namen, tijd en plaats. Naamgeving verschilt aanzienlijk van wijk tot wijk, van regio tot regio. De naamgeving in de stad Utrecht moet bijvoorbeeld op het niveau van postcodewijk bestudeerd worden, willen niet allerlei belangrijke details verloren gaan. Postcodegebieden zijn dan weliswaar ook niet altijd ideale indelingen, maar beter is er niet. Om voldoende gegevens per postcodegebied te hebben, moest de naamgeving over een periode van 20 jaar bijeengenomen worden. Omdat we mogen aannemen dat in die periode de naamgeving een ontwikkeling heeft doorgemaakt, verliezen we het inzicht daarin. Hier werkt de onzekerheidsrelatie: willen we de ontwikkeling in de tijd kunnen beschrijven dan verliezen we inzichten op geografisch niveau, en als we omgekeerd de verschillen in wijken en regio's precies willen beschrijven dan verliezen we het inzicht wat er in de loop van de tijd is gebeurd. In alle gevallen was het nodig om de namen in groepen samen te nemen om hanteerbare en voldoende nauwkeurige tellingen te krijgen. Gelukkig kon dat op zinnige en fraaie manier gerealiseerd worden.

Referenties

Bloothoof, G. (2001). 'Naamgeving binnen Nederlandse gezinnen tussen 1983 en 1999', *Naamkunde* 33, 1-17.

Bloothoof, G., van Nifterick, E. en Gerritzen, D. (2004). 'Over voornamen', *Het Spectrum* (ter perse).

Bloothoof, G., van Nifterick, E. en Gerritzen, D. (2004). 'Van Amalia tot Zjors', *Onze Taal* 73, 60-63.