

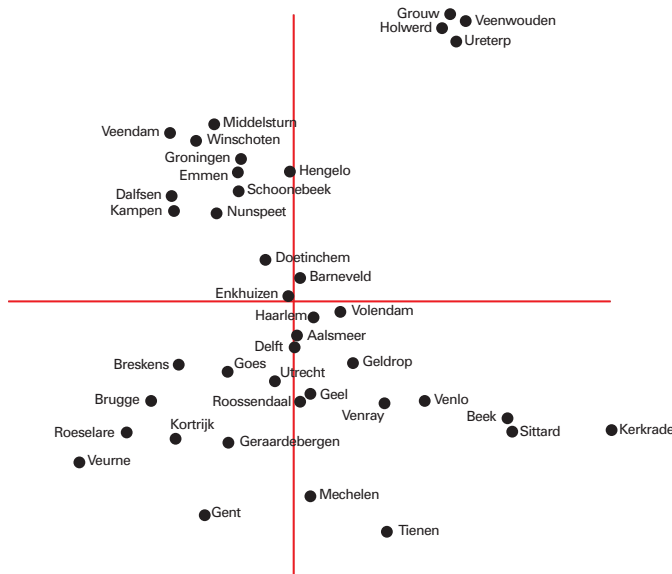
# Universitaire software voor dialectometrie

# De veelkleurige wereld van het

Door: Remco Takken

*In de eerste editie van de Atlas van Nederland die verscheen in de jaren 1963 tot 1977 vinden we naast kaarten die onder andere de variatie in grondsoorten en bodemgesteldheid laten zien ook een dialectkaart. De verschillende kleuren en lijnen laten in één oogopslag de verdeling in verschillende dialectgebieden zien.*

*In de tweede editie is deze kaart niet meer terug te vinden, maar in Groningen zijn wetenschappers nog steeds druk bezig met het in kaart brengen van dialectgebieden.*



Afbeelding 1: Multidimensionale schaling zet sterk verwante dialecten dicht bij elkaar, en sterk verschillende dialecten juist ver uit elkaar. In een dergelijke figuur komen dialectplaatsen soms heel anders ten opzichte van elkaar te liggen dan op een geografische kaart.

De expert op het gebied van dialectcartografie is alfa-informaticus Peter Kleiweg. Hij is als programmeur werkzaam binnen de Groningse Letterenfaculteit en heeft een speciaal softwarepakket ontwikkeld voor het in kaart brengen van dialectvariatie: RUG/L04.

Kleiweg maakte dit pakket RuG/L04 gratis beschikbaar op het Internet en presenteert het daar als een 'programma voor dialectometrie en cartografie'. Dialectometrie is de wetenschap die zich bezighoudt met het meten van taalkundige afstanden tussen dialecten. Dialectometrie gaat aan cartografie vooraf.

Iedere Nederlandse plaats heeft zijn eigen dialect, hoezeer dat ook mag

lijken op de streektaal in een buurdorp. Het dialectlandschap is namelijk een continuüm. Je kunt er als het ware doorheen wandelen en heel geleidelijk de veranderingen waarnemen. Er zijn ook duidelijke grenzen, zoals die tussen de Friese en de Groningse streektalen. Die overgang is veel scherper dan bijvoorbeeld die van het Gronings met het Drents.

### Afstandsmetingen

Om dialectafstanden te kunnen meten, moet een wetenschapper zijn oor eerst te luisteren leggen bij de sprekers van de dialecten. Er worden opnamen gemaakt, en de manier waarop de dialectsprekers woorden uitspreken wordt opgetekend in een

speciaal schrift: het fonetisch alfabet. Met dit alfabet kan heel nauwkeurig worden vastgelegd hoe een woord precies wordt uitgesproken.

Het meten van de verschillen tussen woorden van verschillende dialecten gebeurt met behulp van de Levenshtein-methode. Deze methode werd in 1995 al gebruikt door Brett Kessler die het gebruikte om afstanden te meten tussen Ierse dialecten. Een jaar later paste professor John Nerbonne het samen met een groep studenten toe op Nederlandse dialecten, en dat bleek een groot succes. In de jaren daarna werd de methode onder leiding van Nerbonne verder verfijnd en ook toegepast op Duitse, Noorse, Sardische en Amerikaanse dialecten. De methode kan het best aan de hand van een voorbeeld worden uitgelegd. Het Nederlandse woord 'hart' wordt in het dialect van Vianen net zo uitgesproken als in het standaard Nederlands. Maar in het Belgische Nazareth zegt men 'erte'. Als we de beide woorden op de volgende manier vergelijken:

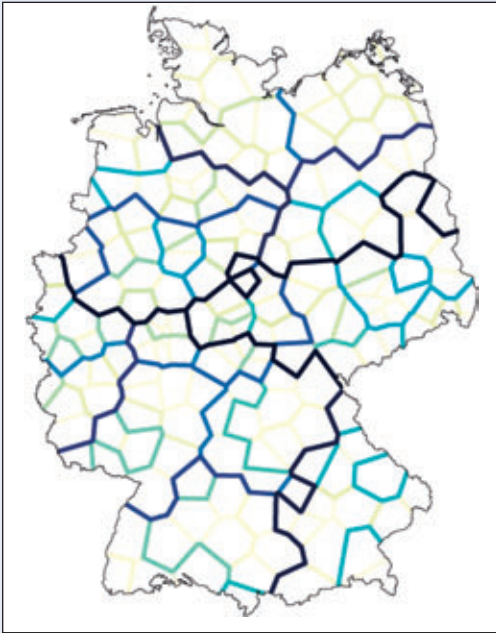
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| h | a | r | t |   |
|   | e | r | t | e |
| * | * |   |   | * |

dan zien we drie verschillen. In Nazareth ontbreekt een 'h' aan het begin terwijl de korte 'a' wordt uitgesproken als een korte 'e'. Verder wordt aan het eind nog een 'e' (maar nu uitgesproken als bijvoorbeeld de 'e' in de) toegevoegd. Er zijn dus drie verschillen. Omdat de woorden een gecombineerde woordlengte hebben van vijf, is de afstand tussen de Viaanse uitspraak en de Nazareense uitspraak gelijk aan  $3/5=0.6$ , oftewel 60 procent.

### Dialectclassificatie

De dialectafstanden die met de Levenshtein-methode worden gemeten, vormen de basis voor de dialectcartografie. Maar voordat een kaart is te maken die de dialectvariatie op een

# dialect in kaart



Afbeelding 2: Dialectvariatie in het Duitse taalgebied op basis van clusteranalyse. Hoe donkerder de lijn, het groter het taalkundig verschil tussen de dialecten aan weerskanten van de lijn. Ontwikkelaar Peter Kleiweg noemt ze 'composite cluster maps'.



Afbeelding 3: Dialectvariatie op basis van clusteranalyse en multidimensionale schaling. Hoe groter het kleurcontrast tussen twee gebieden op de kaart, hoe groter de taalkundige afstand.

begrijpelijke manier laat zien, moeten we de dialecten eerst classificeren. Kleiweg's pakket RUG/L04 biedt twee classificatietechnieken: clusteranalyse en multidimensionale schaling.

Bij clusteranalyse worden dialecten gerangschikt in een boomstructuur op basis van de afstanden die we tussen de dialecten met de Levenshtein-methode hebben gemeten. De boom bestaat uit groepen, groepen kunnen bestaan uit subgroepen, en subgroepen uit subsubgroepen, enzovoorts. Zo ontstaat een hiërarchische boomstructuur. De dialecten vormen in de boom als het ware de bladeren.

Multidimensionale schaling gebruikt dezelfde gegevens als clusteranalyse, namelijk de onderlinge dialectafstanden, maar verder is het een heel andere methode. De dialecten worden zodanig in een twee- of driedimensionale ruimte geplaatst, dat de afstanden die zijn gemeten met de Levenshtein-methode, zo goed mogelijk worden weerspiegeld. Het resultaat is een figuur met stippen en plaatsnaamlabellen die de dialecten representeren. Sterk verwante dialecten zijn in de figuur dicht bij elkaar geplaatst, en sterk verschillende dialecten juist ver uit elkaar, zie afbeelding 1. In een dergelijke figuur komen dialectplaatsen soms heel anders ten opzichte van elkaar te liggen dan op een geografische kaart. Bijvoorbeeld de provincies Groningen en Friesland die geografisch aan elkaar grenzen, liggen in een multidimensionale schaling-figuur een flink eind uit elkaar.

## Dialectcartografie

In de laatste stap worden de classificatieresultaten weergegeven op een kaart. We laten twee kaarten zien die dialectvariatie in het Duitse dialectgebied weergeven. Afbeelding 2 is gebaseerd op de resultaten van clusteranalyse. De kaart geeft in feite de complete boomstructuur weer. Grenzen tussen groepen die in de boom heel significant zijn, worden door donkere lijnen weergegeven, terwijl grenzen tussen minder significante groepen lichter worden weergegeven. Met andere woorden: de grenzen tussen subgroepen zijn

lichter dan tussen groepen, en die tussen subsubgroepen zijn weer lichter dan tussen subgroepen, enzovoorts. Kleiweg heeft deze techniek van weergeven volledig zelf ontwikkeld en noemt dit type kaarten 'composite cluster maps'.

In afbeelding 3 zien we een kaart waarbij clusteranalyse en multidimensionale schaling in combinatie met elkaar zijn gebruikt. Het gecombineerde gebruik van beide technieken is eveneens een vinding van Kleiweg.

In de boomstructuur die we door middel van clusteranalyse krijgen, kunnen we het kortste pad bepalen dat we moeten afleggen om via de takken van het ene blad naar het andere blad te komen. De kortste paden tussen de bladeren, de dialecten, gebruiken we vervolgens als invoer voor multidimensionale schaling. Met deze techniek plaatsen we de dialecten nu in een driedimensionale ruimte. Elk dialect krijgt dus een x-, een y- en een z-coördinaat. Die coördinaten gebruiken we om ieder dialect een kleur te geven. Omdat we drie coördinaten hebben, gebruiken we, net als bij een televisiebeeldbuis, drie kleurcomponenten: rood, groen en blauw. De x-coördinaat bepaalt nu de intensiteit van rood. Bij het dialect met de kleinste x-coördinaat is rood helemaal afwezig, en bij het dialect met de grootste x-coördinaat is rood maximaal aanwezig. Op dezelfde manier bepaalt de y-coördinaat de intensiteit van groen, en de z-coördinaat die van blauw. Daardoor krijgt ieder dialect een unieke kleurmenging en dus een unieke kleur.

We zien dat beide figuren zowel verschillen als overeenkomsten vertonen. Samen geven ze een goed beeld van de dialectvariatie bij onze oosterburen en daarmee beantwoordt RUG/L04 aan het doel waarvoor Kleiweg het pakket heeft ontwikkeld.

Remco Takken

([rtakken@gismagazine.nl](mailto:rtakken@gismagazine.nl)) is redacteur van GIS Magazine. Het pakket RUG/L<sup>04</sup> is te vinden via:

[www.let.rug.nl/~kleiweg/L04/](http://www.let.rug.nl/~kleiweg/L04/).

Daar is ook een uitgebreide handleiding te vinden met meer dialectkaarten.