

Gabmap brengt taalvariaties in kaart

Het beeld en de streektaal

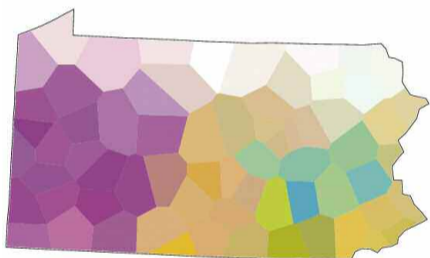
Groningse onderzoekers zijn erin geslaagd taalvariaties te meten en ze overzichtelijk te presenteren.

Erica Renckens

“Alles wat deze nieuwe webapplicatie kan, konden wij technici eigenlijk al. Maar we willen het juist voor anderen óók mogelijk maken. Pas als niet-technisch onderlegde wetenschappers het ook gebruiken is dit project geslaagd.” Aan het woord is John Nerbonne, taaltechnoloog aan de Universiteit Groningen. Tijdens een informatiemiddag van CLARIN geeft hij een presentatie over Gabmap, online software die kwantitatieve analyse van taalvariatie mogelijk maakt. De gebruiker kan met dit programma zelf op een laagdrempelige manier de verschillen tussen taalvarianten in kaart brengen.

Bagage

Onderzoek naar taalvariatie, en met name dialecten, is voornamelijk kwalitatief. De onderzoekers beschrijven de verschillen tussen de taalvarianten, maar kunnen niet nauwkeurig meten in hoeverre deze op verschillende gebieden van elkaar afwijken of juist op elkaar lijken. De software om zulke objectieve vergelijkingen mogelijk te maken bestaat al een tijdje, maar de ge-



In deze Gabmap van Pennsylvania zijn drie variabelen weergegeven. Het blauw/groen duidt op de aanwezigheid van ‘Pennsylvania Dutch’, het wit/grijs op het gebruik van ‘midlands’ dan wel een noordelijke variant van het Amerikaans, en in het paarse gebied spreekt men ‘car’ uit met een /r/, in tegenstelling tot het oranje gebied.



Illustratie Hilje Oosterbaan Martinius

bruiker moet flink wat technische bagage hebben om ermee uit de voeten te kunnen.

“De gebruiker van Gabmap kan zijn data aanbieden in een eenvoudig document waarin de verschillende variabelen met een tab van elkaar gescheiden zijn.” aldus Nerbonne. “Vaak is dit de uitspraak van een bepaald woord, zoals ‘hoes’ en ‘huus’ voor ‘huis’. Maar het kan ook gaan om verschillende woorden voor hetzelfde begrip, zoals ‘jij’, ‘gij’ en ‘du’.”

Kinderliedjes

Gabmap genereert vervolgens automatisch gestructureerde overzichtjes van deze gegevens. De gebruiker kan aangeven van welke variabelen hij de verschillen geanalyseerd wil zien, waarna de software de resultaten van deze analyse ook weer overzichtelijk in kaart brengt, onder andere met behulp van kaartjes uit Google Earth. Op deze kaartjes zie je dan precies welke variant in welk gebied voorkomen. Bovendien is het mogelijk om meerdere variabelen in één kaartje op

te nemen, zodat eventuele verbanden makkelijker zichtbaar worden.

“In de toekomst hopen we ook een vergelijking met andere kaarten mogelijk te maken”, vertelt Nerbonne. “Dan zouden we bijvoorbeeld ook culturele kenmerken in de analyse kunnen meenemen, zoals religie of welke kinderliedjes men in een bepaald gebied zingt.” Maar voorlopig is Nerbonne al erg tevreden met wat de software nu allemaal kan. “We wilden het grootschalig meten van dialectverschillen mogelijk maken, onder andere om de wetenschappelijke discussie over deze totaalaanpak te stimuleren”, verklaart hij. “Het team van Gabmap heeft er hard aan gewerkt om ook de faciliteiten voor de traditionele dialectologie beschikbaar te maken. Dat vergemakkelijkt de vergelijking tussen de kwalitatieve en kwantitatieve methode.”

Gabmap is te gebruiken via www.gabmap.nl. Op deze site zijn ook demonstraties en handleidingen te vinden.

Slagroomtaart voor Friese galgen

Dit jaar reikt SURF foundation elke maand een prijs uit voor de ‘verrijkte publicatie van de maand’. *Peter Boot*

Verrijkte publicaties zijn digitale wetenschappelijke publicaties waaraan niet-traditionele elementen zijn toegevoegd, zoals toegang tot de onderzoeksdata, interactieve illustraties, commentaar of andere bestanden.

In augustus ging de maandelijkse prijs (een slagroomtaart) naar de publicatie *Geen Paniek! Riscocommunicatie door de ogen van de burger*, geschreven door Jan M. Gutting, Marlies Baan, Milou Kievik en Karin Stone van de Universiteit Twente en Deltares. In juli ging de taart naar ‘Gallows in medieval Fri-

sia’, een publicatie die de geografische verspreiding van galgen in middeleeuws Friesland behandelt. Aan de publicatie zijn tabellen, grafieken en een interactieve kaart toegevoegd.

Xpos’re

De galgenpublicatie werd gemaakt met het tool Xpos’re (XML-based Publications On Scientific Research), ontwikkeld door Leen Breure (Universiteit Utrecht). Xpos’re is gebouwd om een demonstratie te kunnen geven van het werk met Geografische Informatie Systemen. Maar de ambities reiken verder, verklaart Breure. “Xpos’re is nu meer een prototype, maar moet een echt product worden, dat vooral voor de auteur gebruiksvriendelijk

is.” Breure wil in het najaar een *requirements study* starten, en ziet de huidige versie van Xpos’re vooral ook als een hulpmiddel om ideeën te concretiseren. Het streven is dat auteurs met enige kennis van ICT zelf met behulp van Xpos’re hun verrijkte publicatie kunnen maken. Maar, zegt Breure, een andere mogelijkheid is dat de verrijkte publicatie gemaakt wordt door bijvoorbeeld bibliotheekmedewerkers of de redactie van een tijdschrift. De bibliotheek van de VU is voornemens een pilot te doen, en Xpos’re zou bijvoorbeeld ook inpasbaar moeten zijn in de workflow die ontwikkeld wordt door het SURF-project ‘Enhanced Journals Made Easy’. Een belangrijke wens is om meer interactieve componenten toe te voe-

gen. Breure: “Op de verlanglijst staan tabellen met filtermogelijkheid en interactieve grafieken. Bij voorkeur maken we daarbij gebruik van bestaande software.” Waarom maakt Xpos’re gebruik van Flash-software? Breure verwijst naar de verschillen tussen browsers (zoals Internet Explorer en Firefox), waarvan hij bij Flash geen last heeft. “Flash is geschikt voor een product waarbij ‘mooi’ en ‘pixel-nauwkeurig’ essentieel is.” Overigens kan Xpos’re ook uitvoer aanmaken voor andere platforms: iPads, smartphones en e-readers.

*www.cs.uu.nl/research/projects/i-cult/xposre/demo/gallows/
www.cs.uu.nl/research/projects/i-cult/xposre/*

KORT

ERIC-status voor ouderenonderzoek

Dit voorjaar heeft de Europese onderzoeksfaciliteit SHARE als eerste de status van een ‘European Research Infrastructure Consortium’ (ERIC) gekregen. De SHARE ERIC is op 25 mei 2011 officieel op het ministerie van OCW geïnstalleerd. SHARE, de ‘Survey of Health, Ageing and Retirement in Europe’ doet in twintig landen langlopend onderzoek naar de bevolking van 50 jaar en ouder. De ERIC is een nieuwe Europese rechtsvorm en is bedoeld om pan-Europese onderzoeksfaciliteiten beter te laten functioneren. Een belangrijk pluspunt is dat de ERIC is vrijgesteld van BTW over diensten en dat scheelt veel geld bij het uitbesteden van het veldwerk. *(Eric Balster)*

www.share-project.org

Breed verlangen naar CLARIAH

Een zeer breed consortium binnen de geesteswetenschappen heeft eind augustus een voorstel ingediend in het kader van de call Nationale Roadmap voor Grootschalige Onderzoeksfaciliteiten van NWO. Het voorstel met de naam CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities) bouwt voort op de Europese projecten CLARIN en DARIAH. Het landelijke consortium bestaat uit alle faculteiten geesteswetenschappen, alle humaniora instituten van de KNAW, het Max Planck Instituut voor Psycholinguïstiek, het Instituut voor Nederlandse Lexicologie, het Nederlandse eScience Center, het instituut voor Beeld en Geluid, de Koninklijke Bibliotheek en het Nationaal Archief. In het CLARIAH project moet een faciliteit voor eHumanities research komen, dat onderzoekers toegang biedt tot een grote hoeveelheid digitale bronnen. *(Ingrid Dillo)*

Betere data met de smartphone

Het Sociaal en Cultureel Planbureau en CentERdata gaan een experiment met smartphones uitvoeren onder leden van het LISS-panel. LISS staat voor Longitudinal Internet Studies for the Social sciences. Het LISS-panel bestaat uit mensen uit alle lagen van de bevolking die via internet vragenlijsten invullen. In het experiment wordt onderzocht of bestaande vragenlijsten met een smartphone kunnen worden afgenomen. Dit is interessant voor bijvoorbeeld het bekende Tijdbestedingsonderzoek (TBO). Als respondenten bereid zijn op hun smartphone een dagboekje bij te houden, kan dat de kwaliteit van de data ten goede komen. Verder wordt bekeken of nieuwe soorten gegevens kunnen worden verzameld: locatiegegevens bijvoorbeeld, of scans van kassabonnen. *(Warna Oosterbaan)*