

# Digitaal prikken in de historie van taal



De digitale wereld dringt de geesteswetenschappen binnen. Letteren is omgedoopt tot e-Humanities en de muis vervangt de boekenlegger. Nicoline van der Sijs verwacht dat de inzet van technologie een schat aan nieuwe kennis zal opleveren voor de historische taalkunde. “De grote vernieuwing zit in de koppeling van bestaande gegevens.”

TEKST: MYRNA TINBERGEN

In een van uw columns in NRC Handelsblad noemt u de mogelijkheden en perspectieven van e-Humanities ‘razend interessant en veelbelovend’. Waarom? “Technologie krijgt een steeds grotere rol binnen de geesteswetenschappen. Niet iedereen is daar even blij mee, maar het is onvermijdelijk. Wetenschappelijke instituten en bibliotheken hebben in de 21ste eeuw massaal digitale bestanden beschikbaar gemaakt. De Digitale Bibliotheek voor de Nederlandse Letteren alleen al bevat vier miljoen pagina’s. Die kun je onmogelijk allemaal geduldig gaan zitten doorlezen. Computers kunnen zo’n databank binnen enkele seconden doorzoeken. Dat levert een stortvloed aan nieuwe gegevens op.”

Sinds 1 januari 2013 bent u hoogleraar ‘Historische taalkunde van het Nederlands in de digitale wereld’ aan de Radboud Universiteit. Dat lijkt paradoxaal: kijken naar taal uit het verleden via het digitale heden.

“Dat lijkt misschien zo, maar mijn proefschrift uit 2001 ging al over etymologie in de digitale wereld. Daarin had ik aan alle woorden uit het Nederlands van nu informatie toegevoegd over wanneer ze voor het eerst voorkwamen, uit welke taal ze afkomstig zijn en uit welk betekenisdomein ze komen. Dat



NICOLINE VAN DER SIJS

maakte het mogelijk allerlei hypothesen te testen. Zo was het algemene idee dat de Nederlanders tijdens de Franse bezetting, rond 1800, uit afschuw van de Fransen geen Franse leenwoorden overnamen. Het tegendeel bleek het geval te zijn. We hebben veel meer hypothesen om te toetsen met digitale technieken. Ik verwacht zeker nieuwe inzichten.”

Zoals?

“Wetenschappelijke ontdekkingen kun je nu eenmaal niet vooraf voorspellen. Maar we weten dat taal verandert en dat immigranten daarbij een rol spelen. Wat is precies de relatie tussen taal en demografische gegevens? Of wat zijn de gevolgen van oorlogen voor taalverandering? Oorlog geeft veel onrust in een samenleving, er zijn dan veel verhuizingen. Welke gevolgen heeft dat voor

dialecten? Dat willen we graag toetsen.”

Dus we gaan de komende jaren veel van u horen?

“Het zal nog wel even duren. In mijn visie zit de grote vernieuwing in de koppeling van bestaande gegevens. Dat moet met beleid gebeuren, en dat kost veel tijd. Hoe snel we dat kunnen doen, hangt ook af van de beschikbare hoeveelheid geld voor onderzoek in de geesteswetenschappen...”

Het nieuwe NWO-project Nederlab, waarvan u projectleider bent, doet onderzoek naar veranderingspatronen in de Nederlandse taal en cultuur. Hoe gebeurt dat?

“Binnen Nederlab kunnen onderzoekers alle digitale teksten van bibliotheken en onderzoeksinstituten tegelijkertijd vanuit één omgeving doorzoeken en analyseren. Ze krijgen zo een enorm bestand aan teksten, van de achtste eeuw tot nu. Daarin kunnen ze bijvoorbeeld onderzoeken hoe de opvattingen over immigranten in de loop der tijd zijn veranderd. Maar we stuiten op allerlei problemen. Sommige teksten zijn gescand en met optische tekenherkenning omgezet naar tekst. Dat gaat voor oudere teksten nog verre van vlekkeloos. Kranten zijn gedrukt op slecht papier, waardoor een ‘s’ soms wordt aangezien voor een ‘f’.

Het woord ‘schaap’ wordt dan weergegeven als ‘fchaap’. We werken nu aan programma’s die zulke leesfouten corrigeren en ook aan slimmere zoekprogramma’s die beter kunnen omgaan met ruis.”

U schreef zo’n dertig (woorden) boeken en tientallen andere publicaties over taal, vooral over het Nederlands. Waarschijnlijk schrijft u net zo snel als u spreekt...

“Haha, kan best. Maar ik heb lange tijd geleefd van de pen en verdiende mijn brood met schrijven. Sinds 2010 ben ik verbonden aan het Meertens Instituut van de Koninklijke Nederlandse Akademie van Wetenschappen. Ik doe onderzoek naar de taalkundige variatie in het Nederlandse taalgebied.”

Kunt u een verrassende ontdekking noemen in uw vakgebied?

“Dat de vorming van onze standaardtaal is beïnvloed door het Duits en niet door het Zuid-Nederlands of Vlaams, zoals algemeen wordt aangenomen. Niet iedereen is het daarmee eens. Toch geef ik in mijn boek *Taal als mensenwerk* uit 2004 sterke aanwijzingen. Met nieuwe onderzoekstechnieken gaat het vast lukken om de sceptici te overtuigen. Zo prik je een beetje in de geschiedenis.” ■