

Besteed de geesteswetenschappen niet uit!

Laatst vertelde een eerstejaars me dat hij in de DBNL bij de dertiende-eeuwse dichteres Hadewych het eerste voorkomen van de term *Nederlands* had gevonden. Ik hoefde de tekst er niet bij te pakken om te weten dat er sprake was van een foute lezing: sinds W.L. de Vreese in 1909 zijn monografie *Over de benamingen onzer taal, inzonderheid over 'Nederlandsch'* publiceerde, heeft een keur aan wetenschappers aandacht besteed aan de vraag hoe oud de benamingen voor onze taal zijn. Een van de opgedane zekerheden is dat *Nederlands* vóór 1480 niet voorkomt: de oudste bron tot nu toe is een Goudse tekst uit 1482. Ook de aanduidingen *Nederland* en *de Nederlanden* voor onze gewesten dateren uit de vijftiende eeuw, en raakten pas echt ingeburgerd toen Karel V in 1548 de Nederlanden in de Bourgondische Kreits verenigde.

Dus nee: Hadewych had het niet over *Nederlands*. De term staat in een verklarende noot die een twintigste-eeuwse redacteur bij de tekst plaatste. Dit eenvoudige voorbeeld toont een interessant probleem aan. Er komen steeds meer moderne en historische teksten digitaal beschikbaar. *Big data* is daardoor een zoekwoord in de geesteswetenschappen: dáár moet de vernieuwing in het onderzoek van komen. Die *big data* kunnen immers met computerprogramma's worden bewerkt en geteld, en dat moet wel tot nieuwe inzichten leiden. In deze context valt vaak het woord *revolutie*. Het probleem is echter dat computers wel allerlei vindplaatsen en cijfers kunnen voortbrengen, maar dat alleen een geschoolde geesteswetenschapper die gegevens adequaat kan interpreteren.

Steeds vaker verschijnen er evenwel artikelen waarin statistici compleet nieuwe correlaties leggen tussen geesteswetenschappelijke data. Daarmee begeven ze zich op glad ijs. Zo beweerde onderzoeker Keith Chen onlangs dat mensen die een taal spreken waarin de toekomstige tijd verplicht wordt uitgedrukt, minder sparen dan mensen waarin dat niet het geval is, zoals het Nederlands, waarin je kunt zeggen *Ik eet morgen thuis*. Mensen die een taal spreken zonder verleden tijd, zouden meer sparen, minder roken en gezonder leven. De werkwoordstijden beïnvloeden de kijk op verleden en toekomst, aldus Chen. Een andere onderzoeker vond een relatie tussen woordvolgorde en kindertal: mensen die een taal spreken met de zinsvolgorde onderwerp, lijdend voorwerp, werkwoord (*zij een kind heeft*) hebben meer kinderen dan mensen die het lijdend voorwerp achter het werkwoord plaatsen (*zij heeft een kind*).

Traditionele geesteswetenschappers hebben dergelijke verbanden nooit gelegd. En daar is een goede reden voor: ze hebben er nooit naar gezocht, omdat de samenhang tussen tijdsuitdrukkingen en spaarzaamheid of kindertal willekeurig is. De computer is een domme telmachine. Hoe meer zaken je telt, hoe vaker je op een toevallige correlatie zult stuiten. Intrigerende voorbeelden staan op de website <http://www.tylervigen.com/>, met onder andere de correlatie tussen het aantal echtscheidingen in Maine en de hoofdelijke consumptie van margarine (advies: eet minder margarine voor een goed huwelijk?).

Een andere fout die makkelijk gemaakt wordt, is overgeneralisatie. Onlangs verscheen een onderzoek naar het plagiaat van Diederik Stapel, waaruit zou blijken dat de computer op basis van het taalgebruik met 70% zekerheid echte artikelen van valse kan onderscheiden. Zelfs al zou dit kloppen (er was direct kritiek op het onderzoek), dan nog bestaat er geen enkele rechtvaardiging van de claim van de onderzoekers dat hun onderzoeksmethode kan worden uitgewerkt tot een tool die valse onderzoeken onderschept. Taalgebruik, zo weet iedere literatuurwetenschapper, is persoonlijk én veranderlijk. Er bestaat geen herkenbaar genre van vals-wetenschappelijke nonfictie – was het leven maar zo simpel...

Hiermee is digitaal onderzoek niet gediskwalificeerd. De genoemde problemen kunnen worden voorkomen als technici en geesteswetenschappers de handen ineenslaan, in plaats van aparte wegen te bewandelen. Sommige onderzoekers, ook geesteswetenschappers, menen dat door de intrede van de computer de eeuwenlang opgedane kennis in de geesteswetenschappen in één klap – revolutionair! - achterhaald raakt. Daarmee gooien ze de baby met het badwater weg. Vernieuwing is veeleer te verwachten van samenwerking waarin geesteswetenschappers en technici gezamenlijk oude vragen tackelen met nieuwe technieken. Geesteswetenschappers moeten daarbij zelf aan het roer staan en de regie niet uitbesteden aan statistici.

Het gedwongen huwelijk tussen technici en geesteswetenschappen betekent, weet ik inmiddels uit ervaring, hard werken aan de relatie. Maar als het huwelijk slaagt, kan het mooie kinderen voortbrengen.