

Tekstanalyticus David Hoover:

Literaire computeranalyse als modern handwerk

HERMAN BRINKMAN

Dit jaar is David L. Hoover, professor of English aan de New York University, te gast in Nederland als visiting KNAW-professor. Hoover is een toonaangevend wetenschapper in de analyse van literaire verschijnselen met behulp van statistische methoden. Begin januari verbleef hij op het Huygens Instituut voor een eerste bezoek, een goede gelegenheid voor een gesprek over zijn werk en de ontwikkelingen in zijn vak.

Voor dat onderzoeksgebied zijn verschillende termen in omloop. Het brede onderzoeksveld wordt wel *digital humanities* genoemd, of *humanities computing*. Een deelgebied daarvan, waar de meeste van Hoovers publicaties zijn te situeren, wordt aangeduid als *computational stylistics*. Daarbinnen heeft *authorship attribution* zijn speciale belangstelling. Auteursonderscheiding (of auteursherkenning) kent een lange geschiedenis. Hoewel de wortels ervan reiken tot in de oudheid, en humanisme en positivisme sterk hebben bijgedragen aan de ontwikkeling van toeschrijvingsmethoden, vond er vooral in de twintigste eeuw een grote doorbraak plaats met de toepassing van statistiek op taalkundige verschijnselen. Later nam die een grote vlucht met de inzet van computerprogramma's en de vorming en beschikbaarstelling van grote tekstcorpora.

Vergelijken

Gevraagd naar de belangrijkste verschillen tussen de meer traditionele methoden en de methoden die hij in de praktijk brengt zegt David Hoover: 'Bij de tradionele methoden zien we vaak minutieuze analyses van één afzonderlijke tekst. Het is echter erg gevaarlijk om je op slechts één auteur te richten, omdat je bij gebrek aan vergelijking geen goede inschatting kunt maken over wat specifiek is voor een bepaalde auteur en wat niet. Het grote verschil met de technieken uit de *digital humanities* is dat er nu enorme tekstcorpora bestaan die dienst kunnen doen als vergelijkingsmateriaal. Vergelijken bestond in het verleden slechts als theoretische mogelijkheid; praktisch gezien was het ondoenlijk. Een ander belangrijk verschil is de houding ten opzichte van de bewijsvoering. Bij de traditionele methoden, die overigens nog altijd worden beoefend, is die voor een groot deel retorisch, op overtuiging gericht, terwijl de bewijsvoering in de *digital humanities* veel meer vraagt van de kwantitatieve aspecten van de bewijslast.'

Argwaan

Ondanks het feit dat er in de laatste jaren opmerkelijke resultaten zijn geboekt, kent nog maar een beperkt deel van de onderzoekers in de geesteswetenschappen de



nieuwe methoden een plaats toe in het onderzoek. Volgens Hoover zijn daar verschillende redenen voor. Allereerst heeft dat te maken met de wijze waarop in de *humanities* wordt gekeken naar meer empirisch gefundeerde manieren van kennisverwerving. Er is volgens hem nog altijd argwaan tegenover een positivistische, op bewijzen gegrondveste tekstanalyse binnen de menswetenschappen.

Veel wetenschappers vrezen aangezien te worden voor handswerksman in plaats van theoreticus. Het aanzien van de theorie blijft het allerhoogste'

Tevens vermoedt hij de invloed van het postmodernisme, *new historicism* en context-gerelateerde benaderingen, waarbij het door de onderzoeker gehanteerde concept dikwijls belangrijker wordt geacht dan het onderzoeksobject. 'Er is over het algemeen nog steeds weinig interesse voor onderzoek dat louter tekstueel is gericht.'

Postkoloniale vooroordelen

De universeel toepasbare analysemodellen liggen bovendien klaar om te worden toegepast. 'Met een hamer in de hand voelt alles om je heen aan als een spijker. Ik heb meegemaakt dat een onderzoekster de stelling wilde verdedigen dat Virginia Woolfs *To the Lighthouse* uit 1927 gebaseerd was op postkoloniale vooroordelen. De enige feitelijke bewijsgrond die ze had was

dat over een van de personages ik meen vijf keer werd gezegd dat ze met oriëntaalse spleetogjes ergens naar keek. En dat terwijl die roman alleen gaat over een groep blanke mensen die zich ophouden aan de Schotse kust. Het gaat me niet om dit specifieke voorbeeld, maar om het feit dat teksten niet zo 'open' zijn als de postmodernisten graag zouden willen. Tekstinterpretatie zal zijn fundament toch moeten vinden in de tekst zelf; de tekst zelf schept de grenzen van de interpretatie ervan.'

Een concreet voorbeeld van de manier waarop *computational stylistics* een toegevoegde en corrigerende waarde heeft zie je bij analyse van het werk van Henry James. 'We weten dat hij op zekere leeftijd vanwege rheumatische problemen zijn romans ging dicteren in plaats van ze zelf te schrijven. Daarbij is er al vaak geconstateerd dat er een aanzienlijk verschil bestaat tussen zijn vroege en late stijl.

Nu hebben sommigen beweerd dat die stijlontwikkeling is toe te schrijven aan de overgang van schrijven naar dicteren. We kunnen het niet bewijzen aan de hand van manuscripten, want die zijn er niet. Maar als we nu een digitale stijlanalyse maken van zijn romans, blijkt er nergens een scherpe overgang. Wel kunnen we een duidelijke ontwikkeling laten zien, maar die is overal gradueel. Dat betekent dat James stijlontwikkeling geheel losstaat van de manier waarop hij zijn romans schreef.'

Vruchtbaar samengaan

Hoover is niet van mening dat de corpus-gerelateerde onderzoeksmethoden de traditionele, niet-statistische analysemodellen overbodig maken. Er is alle ruimte voor een vruchtbaar samengaan, met name wanneer het gaat om de vervolgvraag. 'Zoals Hugh Craig (professor aan de University of Newcastle, Australia - HB) zegt: "Als je eindelijk zekerheid hebt over het auteurschap, wat ben je dan eigenlijk te weten gekomen, wat is het precies dat je hebt geleerd?"'

Politieke lading

De invloed van de *digital humanities* blijft, vreemd genoeg, achter bij de successen die er duidelijk worden geboekt. Anders dan je misschien zou denken is het in de Verenigde Staten, maar ook elders in de wereld, niet eenvoudig om voor projecten in de *digital humanities* onderzoeksfianciering te krijgen. Hoe valt dat te verklaren? 'Niet iedereen zit te wachten op onderzoek dat mooie

theorieën ontkracht. Sommige theorieën hebben soms ook een politieke lading. Zo doe ik op het ogenblik onderzoek naar een anonieme tekst waarvan wel wordt beweerd dat hij door Edgar Allen Poe is geschreven. Als dat zo zou zijn, zou hij te boek komen te staan als een 'race theorist'. Zoals het er nu naar uitziet is Poe geenszins de auteur van die tekst, maar sommigen hebben moeite om dat te accepteren.'

Teveel theoretiseren

Dikwijls zijn de verwachtingen over de uitkomsten van onderzoeken ook te hoog gespannen. Men verlangt, zeker als de computer wordt ingezet, een ondubbelzinnig oordeel maar een oordeel kan soms niet anders dan onder het nodige voorbehoud worden gegeven.

'Vergelijken bestond in het verleden slechts als theoretische mogelijkheid; praktisch gezien was het ondoenlijk'

Hoovers indruk van het huidige debat onder zijn collega's, met name onder degenen die werken met computeranalyses, is dat men over het algemeen teveel blijft steken in het theoretiseren over wat er zou moeten gebeuren in plaats van het implementeren van wat er al mogelijk is. 'Veel wetenschappers vrezen het nu eenmaal aangezien te worden voor een handswerksman in plaats van een theoreticus. Hetaanzien van de theorie blijft het allerhoogste.'

Historische letterkunde

Waar de *digital humanities* echt voet aan de grond hebben gekregen, in het bijzonder doordat er duidelijke resultaten zijn geboekt, is het gebied van de historische letterkunde en met name de achttiende en negentiende eeuw. Ook in mediëvistische kringen groeit de belangstelling. 'Misschien komt dat doordat bij onderzoekers in deze specialismen er altijd al wel een besef is geweest dat men hoe dan ook op teksten gericht moet blijven. Daar ligt de kern van ons vakgebied.'

David Hoover zal in mei en juni van dit jaar terugkeren naar Nederland voor een tweede en derde termijn van gastaanstelling.

David L. Hoover is professor in de Engelse Taal aan de Universiteit van New York, waar hij sinds 1981 les geeft. Hij haalde zijn bachelors in Engels en Filosofie in 1971 aan het Manchester College en zijn masters in 1974 aan de Universiteit van Indiana, waar hij in 1980 promoveerde op een proefschrift over het oud-Engelse dichtmetrum. Zijn onderzoek richt zich behalve op dat metrum op digital humanities, computational stylistics, corpus stylistics, auteursherkenning. Recente boeken: *Stylistics: Prospect and Retrospect* (2007) en *Language and Style in The Inheritors* (1999).

