

## Literatuurwetenschap zonder hoogtevrees



Illustratie: Vince Trommel.

In de rubriek ‘Verstand van Zaken’ reflecteren deskundigen in een essayistische vorm op de stand van zaken in de hedendaagse literatuur, het boekenvak of de wetenschap. Trends worden geduid, catastrofes voorzien en normen bevroegd. Hoogleraar Computationele literatuurwetenschap Karina van Dalen-Oskam geeft een overzicht van het onderzoek op het gebied van digital humanities in binnen- en buitenland. Ze reflecteert op de controverses en discussies binnen het vakgebied. Daarnaast gaat ze dieper in op ‘The riddle of literary quality’, een groots opgezet computationeel onderzoeksproject met als onderwerp de hedendaagse percepties van literariteit in Nederland. Het onderzoek vormde de basis voor meerdere publicaties in het veld van computationele literatuurwetenschap. Zijn computationele middelen een toevoeging aan de gereedschapskist van de literatuurwetenschapper of kunnen ze beter geweerd worden?

### Distant reading

In het dit jaar verschenen boek *Distant Horizons* schrijft de bekende computationele literatuurwetenschapper Ted Underwood dat er nog steeds heel wat angst bestaat voor de combinatie literatuur en computer. Hij doet zijn best om de zorgen weg te nemen over *distant reading* – het met hulp van software analyseren van meer data dan we ooit als mens konden lezen.

Underwood doet onderzoek naar Amerikaanse literaire ontwikkelingen over een lange periode. Hij betoogt dat de literatuurgeschiedenis tot nu toe beschouwd is als een opeenvolging van herkenbare literaire perioden, simpelweg omdat het brein van een onderzoeker geen grotere porties tegelijkertijd kan onthouden en analyseren. De schaal van de langetermijntontwikkeling maakt dus dat slechts één mens dat verloop niet kan doorgronden. ‘A single pair of eyes at ground level can’t grasp the curve of the horizon, and arguments limited by a single reader’s memory can’t reveal the largest patterns organizing literary history.’ (Underwood 2019a: ix-x) Maar de machine kan die schaal wel aan – tenminste, als we die verstandig bedienen. We moeten gewoon meer afstand nemen. De lucht in, om het grotere plaatje te kunnen zien. Daar hoeven we niet bang voor te zijn: het is een kwestie van literatuurwetenschap zonder hoogtevrees.

### Oorlog in de geesteswetenschappen

Vanuit Nederlands perspectief is het verrassend dat Underwood zoveel nadruk legt

op het wegnemen van angst – in Nederland lijken de nieuwe methoden en technieken in de geesteswetenschappen al behoorlijk ingeburgerd te raken. Zo niet in Noord-Amerika. *The chronicle of higher education* spreekt zelfs van ‘The Digital Humanities Wars’. Op 27 maart 2019 verscheen daarin een artikel van Nan Z. Da, die stelt dat computationele methoden in de digitale geesteswetenschappen vaak nog danig tekort schieten.

De opzet van Da’s artikel is lovenswaardig: ze probeert het onderzoek van Underwood en anderen te repliceren. Replicatiestudies zijn een nieuwe en belangrijke ontwikkeling in geesteswetenschappelijk onderzoek na de *digital turn*. Door dezelfde teksten en methoden te gebruiken, met de softwarecode zoals die door de makers beschikbaar zijn gesteld, gaat Da na of ze tot dezelfde meetresultaten en conclusies komt. Het resultaat: Da is nog niet overtuigd van de meerwaarde van computationele methoden. Ze zijn minder innovatief en complex dan de retoriek van de onderzoekers doet geloven. In een stuk in *Critical Enquiry* is Da nog strenger: het verkennen van nieuwe methoden en technieken is op zich goed, maar het is nu tijd om te kijken of het al het geld dat erin is gestopt waard is. Moeten we hier wel mee verdergaan?

### **Computationele literatuurwetenschap in Nederland**

De moederen lijken minder hoog op te lopen in Nederland. In de letterkundige mediëvistiek zijn de nieuwe digitale en computationele methoden en technieken al enige tijd geleden omarmd.<sup>1</sup> Langzamerhand begint ook de moderne letterkunde ze te verkennen. In 2012 werd de eerste (deeltijds)leerstoel Computationele literatuurwetenschap opgericht aan de Universiteit van Amsterdam. Daaruit voortkomend wordt sinds 2014 door de Onderzoeksschool Literatuurwetenschap jaarlijks een introductie cursus Computational Literary Studies aangeboden die door al heel wat promovendi en studenten van verschillende universiteiten is gevolgd. Momenteel werken verschillende promovendi aan dissertaties met een computationele invalshoek, zoals Lucas van der Deijl (UvA) en Roel Smeets (RU), die samen met Saskia Pieterse en Marion Prinse ook verantwoordelijk zijn voor de bekende Personagebank. Aan het Huygens ING is in 2018 een project gestart waarin het digitale schrijfproces van auteurs met nieuwe middelen wordt onderzocht door promovendi Floor Buschenhenke en Lamyk Bekius. Daarnaast zijn er meer promotietrajecten en projecten waarin het computationele aspect een rol speelt.

### **The Riddle of Literary Quality**

Het allereerste computationele onderzoeksproject rond moderne literatuur, *The Riddle of Literary Quality*, verkreeg in 2011 subsidie uit het Computational Humanities Programme van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW). Het ging in 2012 van start. Els Stronks beschreef het in 2013 in *Vooy* in haar

1 Zie voor een historisch overzicht van de ontwikkeling van de digitale geesteswetenschappen in Nederland Van Zundert en Van Dalen-Oskam 2014.

bijdrage aan deze reeks: ‘In het project wordt de relatie tussen tekstimmanente kenmerken en de waardering van lezers onderzocht. Welke teksten zien lezers als literair, en correspondeert dit met bepaalde tekstkenmerken (complexiteit van zinnen, taalgebruik et cetera) die met behulp van computerondersteunde analyses van die teksten vastgesteld kunnen worden?’ (Stronks 2013: 79-80) Stronks riep haar lezers op om in de gaten te houden welke resultaten dit project met een ‘wrijving van inzichten’ (82) op gaat leveren.

Het project, dat in december 2019 officieel wordt afgerond, maakt gebruik van een nieuwe mix van methoden om uit te vinden wat de hedendaagse percepties van literariteit in Nederland zijn. Ons team onderzocht zowel de sociologische kant als de tekstuele kant. In 2013 verzamelden we in Het Nationale Lezersonderzoek lezersmeningen over 401 recente romans die het meest verkocht of uitgeleend waren in de drie jaar daarvoor. We wilden weten welke boeken de deelnemers hadden gelezen en hoe ze een aantal van de gelezen en niet gelezen boeken waardeerden. Ze gaven een score op twee schalen: literaire kwaliteit en algemene kwaliteit. Bijna 14.000 mensen deden mee, waaronder veel ‘gewone’ lezers, maar ook lezers die professioneel met boeken bezig zijn, zoals onderzoekers of recensenten. We constateerden dat lezers anders oordeelden over romans uit verschillende genres: literaire fictie scoorde het hoogst, daarna spannende boeken en vervolgens romans met het label romantiek. We zagen hoe de selectiecriteria voor ons corpus van romans evenveel boeken geschreven door vrouwen als door mannen hadden opgeleverd, maar dat de verdeling per genre aanzienlijk verschilde. Romantiek was voornamelijk geschreven door vrouwen en bij de ‘literaire romans’ overwoog het aandeel van mannelijke auteurs. Wat de gemiddelde scores per boek betreft, viel vooral op dat in de categorie literaire roman boeken van vrouwelijke auteurs significant lager werden gewaardeerd dan die van mannelijke auteurs. Verder kregen uit het Engels vertaalde literaire romans een wat lagere score op literariteit dan in het Nederlands geschreven romans.

### **Is literariteit meetbaar?**

De andere poot van het onderzoek had niet de lezers maar de teksten zelf als object. We gingen na of de boeken waarover onze lezers het min of meer eens waren dat ze niet erg literair waren inderdaad anders waren dan de boeken die wel een hoge score voor literariteit kregen. Dat deden we door de romans met hulp van software door te meten. De metingen betroffen woorden, woordsoorten, woordvariatie, complexiteit van zinsstructuur en gebruik van clichés, om alleen de belangrijkste talige aspecten te noemen (er waren er nog veel meer).

Het eerste uit het project voortgekomen proefschrift, *Rich statistical parsing and literary language* (2016) van Andreas van Cranenburgh, concentreert zich op deze metingen. Van Cranenburgh heeft nieuwe manieren ontwikkeld om patronen in woordgebruik en zinsbouw van de romans op te sporen. Zijn onderzoek laat zien dat romans die verschillend worden gewaardeerd door de lezers ook meetbaar in taalgebruik van elkaar verschillen. Literaire romans hebben bijvoorbeeld minder clichés dan spannende

en romantische boeken en gaan over meer verschillende onderwerpen. Boeken die in hoge mate literair werden gevonden hadden meestal (maar niet altijd) een grotere woordenschat dan minder literaire boeken en bevatten meestal (maar niet altijd) gemiddeld langere zinnen. Van Cranenburgh ontwikkelde verder een model waarmee lezersmeningen op basis van tekstkenmerken voorspeld kunnen worden. Het ligt voor de hand dat er meer speelt dan de tekst, het genre en het auteursgeslacht. Juist de boeken waarvoor het model een afwijkende score gaf van die van de lezers zijn interessant. Van Cranenburghs model schatte bijvoorbeeld *De afrekening* van Mart Smeets aanzienlijk hoger in dan onze lezers hadden gedaan, net als de roman *Eten, bidden, beminnen* van Elizabeth Gilbert. En de literaire nummer één voor de lezers, *Alsof het voorbij is* van Julian Barnes, kreeg juist een aanzienlijk lagere score. Onze hypothese is dat die afwijkingen waarschijnlijk te maken hebben met de invloed van allerlei vormen van prestige – en dat we dat door onze metingen kunnen laten zien is een van de grote methodologische innovaties uit ons project.

### **Vooroordelen**

Uit Het Nationale Lezersonderzoek bleek dat de deelnemers (waarvan 71 procent trouwens vrouw was) een hogere literariteit toeschreven aan werken van mannelijke auteurs dan aan werken van vrouwelijke auteurs. (Koolen 2018: 68-76) Dit opvallende verschil werd het onderwerp van het tweede proefschrift dat uit het project is voortgekomen, *Reading beyond the female. The relationship between perception of author gender and literary quality* (2018) van Corina Koolen. Dit proefschrift laat onder meer zien dat lezers ook in hun motivaties boeken van vrouwelijke auteurs anders benaderen dan die van mannelijke auteurs. In een vervolgonderzoek vergeleek Koolen een gelijke hoeveelheid romans van vrouwen en mannen die zijn genomineerd voor een literaire prijs, en ook daar komen observaties uit voort die het onderzoek in een nieuwe richting sturen. (Koolen 2018: 128-161) Dit proefschrift heeft heel wat media-aandacht gekregen en Corina Koolen heeft zo een belangrijke bijdrage kunnen leveren aan de weerkerende discussie over het aandeel van vrouwen in de hedendaagse Nederlandse literatuur. Koolen werkt haar Engelstalige proefschrift momenteel om naar een Nederlands boek voor een breed publiek.

### **De toekomst van de computationele literatuurwetenschap**

Op dezelfde dag dat Nan Z. Da's kritische stuk in *The chronicle of higher education* verscheen, werd daar een reactie gepubliceerd van de hand van Ted Underwood. De titel vat zijn antwoord op Da's aanval samen: 'Dear humanists: fear not the digital revolution. Advances in computation will benefit traditional scholarship – not compete with it.' Underwood maakt nogmaals duidelijk dat er niets te vrezen is. De literatuurwetenschap heeft er alleen maar mogelijkheden bijgekregen. Het zou bijzonder onverstandig zijn om nieuwe methoden te weren omdat ze zo ongewoon zijn in het onderzoek tot nu toe. Bovendien, zo stelt Underwood, zijn heel veel wetenschappers uit andere vakgebieden razend benieuwd naar de mogelijkheden, zozeer zelfs dat ze de nieuwe

methoden ook zelf al gaan toepassen en daar heel wat media-aandacht mee genereren. In plaats van de nieuwe mogelijkheden volledig over te laten aan niet-geesteswetenschappers kunnen we er maar beter zelf ook mee aan de slag gaan, gebruikmakend van al bestaande en bewezen ‘traditionele’ onderzoeksmethoden. We kunnen zo verder bouwen op de kennis die de literatuurwetenschap al verzameld heeft en in een combinatie van bestaande en nieuwe methoden onze kennis op een verantwoorde manier uitbreiden. Er is tenslotte nog heel erg veel te onderzoeken.

Het zal niet verbazen dat ik het eens ben met Underwood. Computationale methoden zijn niet meer weg te denken uit de literatuurwetenschap. Ze breiden de gereedschapskist van de onderzoeker uit en leiden tot veranderingen in de manier van werken en denken. Ook de opleidingen veranderen. Studenten worden voorbereid op een literatuurwetenschap met computationale middelen. Met die nieuwe kennis krijgen toekomstige onderzoekers de keuze om in meerdere of in mindere mate van die mogelijkheden gebruik te maken. Ik verwacht dat steeds meer onderzoekers met plezier en profijt deze *distant horizons* gaan verkennen. Zonder hoogtevrees.

‘*Computationale methoden zijn niet meer weg te denken uit de literatuurwetenschap.*’

## Literatuur

- Van Cranenburgh, A.**, *Rich statistical parsing and literary language*, Amsterdam 2016, <<https://hdl.handle.net/11245/1.543163>>.
- Da, N.Z.**, ‘The computational case against Computational Literary Studies’. In: *Critical Inquiry* 4 (2019): 601-639.
- Da, N.Z.**, ‘The Digital Humanities debacle. Computational methods repeatedly come up short’. In: *The Chronicle of Higher Education*, <<https://www.chronicle.com/article/The-Digital-Humanities-Debate/245986#>>, laatst geraadpleegd: 08-07-2019.
- Dalen-Oskam, K. van**, ‘The Riddle of Literary Quality. Op zoek naar conventies van literariteit’. In: *Vooy's* 32 (2014) 3: 25-33.
- Koolen, C.**, *Reading beyond the female. The relationship between perception of author gender and literary quality*, Amsterdam 2018, <<https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37fice7>>.
- Stronks, E.**, ‘Lezen en rekenen’. In: *Vooy's* 31 (2013) 3/4: 78-82.
- Underwood, T.**, *Distant horizons. Digital evidence and literary change*, Chicago/London 2019a.
- Underwood, T.**, ‘Dear Humanists: Fear Not the Digital Revolution. Advances in computation will benefit traditional scholarship – not compete with it’. In: *The Chronicle of Higher Education*, <<https://www.chronicle.com/article/Dear-Humanists-Fear-Not-the/245987>>, laatst geraadpleegd: 08-07-2019.
- Zundert, J. van & K. van Dalen-Oskam**, ‘Digital Humanities in the Netherlands’. In: *H-Soz-Kult: Kommunikation und Fachinformation für die Geschichtswissenschaften*, <<https://www.hsozkult.de/debate/id/diskussionen-2396>>, laatst geraadpleegd: 08-07-2019.