

Els Stronks

Lezen en rekenen

In de nieuwe 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 bevraagd. Hoe kan digitalisering een toevoeging bieden aan geesteswetenschappelijk onderzoek? Hoogleraar Els Stronks geeft in de 'Verstand van Zaken' van dit nummer hierover aanzet tot nadenken. Ook geeft zij voorbeelden van onderzoek waarin digitalisering van nut kan zijn, en legt zij uit hoe digitalisering ervoor kan zorgen dat wij met herontdekte methodes teksten kunnen bestuderen. Zij legt uit hoe door middel van wederzijdse kennis informatiekundigen en geesteswetenschappers elkaars werk kunnen versterken.

De stand van zaken

De letterkundige kan goed lezen, de computer goed rekenen: het is al decennialang de vraag hoe die twee sterktes elkaar kunnen vinden. Om het wat gechargeerd te stellen: informatici zien literatuur als een corpus teksten waarvan de digitale ontsluiting en modellering aangepakt kunnen worden naar analogie met methoden die voor andere datasets ontwikkeld worden. In hun vak wordt groter inzicht in die datasets verkregen doordat een abstractie van die data (een model) gemaakt wordt, waarmee vervolgens wordt gerekend. In die abstractie is het mogelijk patronen te vinden die onbewerkte of ongemodelleerde datasets niet goed waarneembaar zijn en die tot groter inzicht in de data leiden. Literatuurwetenschappers zien literatuur als een domein dat zich weet te onttrekken aan wetten van logica en systematiek die voor andere data geldig zijn. Literaire teksten zijn grillig, afwijkend en bewust ontregelend. Het is daarom moeilijk voor te stellen dat met rekenkracht en een abstract model meer inzicht in literatuur wordt verkregen dan letterkundigen zelf al lezend kunnen produceren.

Toch lukt dat, met name op het terrein van de stylometrie, de stijlleer. Door in literaire teksten functiewoorden te tellen (lidwoorden, tussenwerpsels en voegwoorden, die samen dus het abstracte model vormen waarmee stijl wordt ontleed), kan de specifieke stijl van een auteur of zelfs de specifieke stijl van een genre gedetecteerd worden. Zo wist Mike Kerstemont het auteurschap van enkele anonieme Middelnederlandse teksten met behulp van door de computer uitgevoerde stijlanalyses overtuigend vast te stellen. (Van Zundert 2012)

Maar de stylometrie is maar een onderdeel van de letterkunde, kan de rekenkracht van de computer ook in andere typen letterkundig onderzoek ingezet worden? In een



recent gestart KNAW-project ‘Riddle of Literary Quality’ van Karina van Dalen-Oskam en Rens Bod wordt daartoe een poging gedaan. In het project lijken letterkunde en informatica diametraal tegenover elkaar gepositioneerd voor een maximaal effect. Basis van het project is de onderzoeksvraag: ‘wat is literatuur?’. Die vraag is door wetenschappers, in het spoor van Bourdieu en anderen, vaak beantwoord als ‘dat wat instituties ervan maken’. Recensenten, uitgevers en media zijn dan de instituties die bepalen waar literaire kwaliteit uit bestaat. Dat oordeel is tijdsgebonden, en kan door instituties ter discussie worden gesteld. Die vraag leek daarmee afdoende beantwoord (concludeerde taalkundige Marc van Oostendorp onlangs (Van Oostendorp 2013), maar in het project ‘Riddle of Literary Quality’ wordt de kwestie aangekaart om te zien of met behulp van de computer een ander antwoord kan worden gegenereerd. 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? Het project vertrekt vanuit een vraag die uit de literatuurwetenschap komt, loopt verder in een spoor dat de literatuurwetenschap vreemd is maar wel eigen is aan de computationele taalkunde (een geesteswetenschappelijke discipline waarin het modelmatig analyseren van de informatici veel weerklank heeft gevonden).

*‘Het is moeilijk voor te stellen
dat met rekenkracht meer
inzicht in literatuur wordt
verkregen dan letterkundigen
zelf al lezend kunnen
produceren.’*

Het kan ook nog anders, de computer en letterkunde combineren, laat het artikel 'How Literature Becomes Knowledge: A Case Study' van de Amerikaanse onderzoekster Robin Valenza uit 2009 zien. Het artikel biedt een mix van drie onderzoeksgebieden die mij inhoudelijk interesseren (historisch leesgedrag, de productie van kennis in de vroegmoderne tekstuele cultuur en digitalisering), maar in algemenere zin is interessant hoe Valenza opereert tussen Literatuurwetenschap en Informatica. Ze bekijkt digitaliseringsvraagstukken en literaire cultuur niet vanuit puur informatiewetenschappelijk dan wel literatuurwetenschappelijk perspectief, maar zoekt naar raakvlakken en overeenkomsten die inzicht kunnen bieden.

Valenza bekijkt daartoe in haar artikel de index die kort na het verschijnen van Samuel Richardsons *Clarissa* in 1748 aan de roman toegevoegd werd. Voor *Clarissa* bestond al meteen veel belangstelling, maar de omvang van de in totaal zeven delen tellende roman vormde een probleem. Een moderne Penguin-uitgave van de gehele tekst omvat zo'n 1500 pagina's, om een indruk te geven van de hoeveelheid pagina's waar achttiende-eeuwse lezers zich doorheen moesten zien te werken. Het is niet de verhaallijn die zoveel tekst vergt, want die kan in de flaptekst van de Penguin-uitgave in één zin worden samengevat: 'Clarissa, in resisting parental pressure to marry a loathsome man for his money, falls prey to Lovelace, is raped and dies.' Het zijn de *sentiments* (reflecties op de handeling van alle personages) die de roman zo omvangrijk maken en die ook de meeste aantrekkingskracht op de lezers uitoefenden.

De *index rerum* (index op onderwerp) zou die reflecties toegankelijk moeten maken en dat bleek in de praktijk ook zo te werken. Lezers vonden er alle passages over bepaalde onderwerpen handig bij elkaar staan. Zo'n onderwerp was bijvoorbeeld 'Advices and Cautions to Women', met verwijzingen naar tachtig kort getypeerde passages, zoals: 'Every one's eyes are upon the conduct, the visits, and the visitors of a young Lady made early independent, i. 120 [=Volume I, p. 120]'. Het was Richardsons uitgever die op het idee kwam die indexen toe te voegen, en Richardson zelf die voor de uitvoering ervan in 1751 zorg droeg.

Valenza stelt aan de hand van deze casus een vraag die uit de hoek van historisch leesonderzoek komt: vormden die indexen geen uitnodiging voor de vroegmoderne lezer om oppervlakkig te lezen? De nuances en tegenstrijdigheden uit de verhaallijn en karaktertekening worden weggevaagd door een indexerend systeem dat dwingt passages te vergelijken, te ordenen, te beoordelen en aan elkaar gelijk te verklaren. Een systeem dat de lezer een trage gang langs de verhaallijn lijkt te ontraden. Die indexen stimuleren, zo betoogt Valenza, dus een andere manier van lezen. Die manier van lezen wordt gefaciliteerd door de indexen, die op moderne databases lijken. Want ook daarin wordt geabstraheerd en gegeneraliseerd. Zoals Richardson passages bij elkaar zette die naar zijn idee onder dezelfde noemer vielen, zo ordent de maker de onderliggende data in een database volgens een overzichtelijke structuur.

Dat type ordenen van gegevens, dat zo modern lijkt, heeft dus een eeuwenoude geschiedenis. Want Richardsons *Clarissa* (met index) is slechts een van de vele historische teksten die zo ontsloten is. Het principe achter deze vorm van ontsluiten bestond

al veel eerder en was wijdverbreid. Het had de functie teksten van allerlei soort – ook teksten die wij in de loop der tijden literair zijn gaan noemen – handige bronnen van kennis te maken. De index stimuleert dus niet tot oppervlakkig lezen, maar beoogt de kennis van de lezer te vergroten.

Valenza typeert het indexerend lezen in de *Clarissa* als een historische vorm van *distant reading*. Die term is jaren geleden bedacht door de literatuurwetenschapper Franco Moretti. (zie onder anderen Moretti 2005) Moretti is inmiddels werkzaam in het door Silicon Valley gesponsorde *literary lab* in Stanford: wat hij aan analysemethode verzong, wordt momenteel door velen gewaardeerd. Moretti introduceerde die term in 2000 op polemische toon: ‘we weten al hoe we moeten lezen’, hield hij letterkundigen voor, ‘laten we nu leren hoe niet te lezen’. Hij bedoelde daarmee wat in de stylometrie nu goed is gelukt: door op functiewoorden te letten, de woorden waaraan letterkundigen traditioneel veel minder betekenis hechten dan aan zelfstandig naamwoorden, werkwoorden et cetera, worden patronen in stijl zichtbaar. Het indexerend lezen zou je hiervan een verre voorganger kunnen noemen.

Valenza’s betoog voert haar verder naar het moment waarop *close reading* ontstond. Als letterkundigen hebben wij weinig waardering voor indexerend lezen en hebben de indexen bij *Clarissa* daarom lang genegeerd of afgedaan als hulpmiddelen voor lezers die de hele roman lezen niet aan zouden kunnen – als tekenen dus van inferieur lezen. Dat, zo betoogt Valenza, is te wijten aan de New Critics. Die stimuleerden letterkundigen niet alleen tot *close reading* van literaire teksten, maar droegen ook uit dat literaire teksten beschouwd moeten worden als op zichzelf staande objecten, en niet als dragers van kennis over de wereld buiten de literatuur. ‘Literary critical reading a la the New Critics might thus be defined as the particular form of reading that opposes index-learning.’ (Valenza 2009: 226) De New Critics waren in hun promotie van dat *close reading* zo succesvol, dat hun uitgangspunt van *close reading* inmiddels als een soort juk op de letterkundigen rust, en historische vormen van *distant reading* aan het zicht werden onttrokken.

Valenza’s onderzoek interesseert mij omdat ze een kritische blik werpt op de letterkundige onderzoekspraktijk, en met name op het belang van het *close reading*-ideaal. Maar wat me ook interesseert is de relatie die ze legt tussen processen van kennisoverdracht en literaire teksten. Het type index dat Richardson voor *Clarissa* maakte, was aangepast aan het snelle, indexerende lezen dat toegang tot op onderwerp geordende kenniscomplexen bood. Zit hier misschien een interessante invalshoek voor grootschaliger computerondersteund onderzoek naar het vermogen van literatuur om complexe kennis aan lezers over te dragen?

Deze specifieke inhoudelijke vraag even latend voor wat hij is, wil ik nu kort bekijken hoe vanuit hier verder te gaan met de zoektocht naar een zinvolle en productieve combinatie van computer en lezer. Valenza ageert nogal tegen de New Critics. In een ander artikel van de drie ‘digital humanities specialists’ Gooding, Terras en Warwick worden niet de New Critics, maar de menselijke angst voor nieuwe technologieën als oorzaak voor de geringe toenadering tussen letterkundigen en informatici genoemd.

(Gooding e.a. 2013) Wie of wat er ook precies verantwoordelijk is voor de huidige situatie, ik denk dat we als letterkundigen meer verstand van informatica zullen moeten krijgen om progressie te maken. Pas dan kunnen we zelf actief suggereren hoe dat vakgebied ons kan helpen.

Omdat ik dat denk, ben ik blij dat we sinds september in Utrecht een minor Digital Humanities voor BA-studenten uit de faculteit Geesteswetenschappen hebben waarin Informatica een duidelijke plaats en functie heeft gekregen. We gaan in de minor niet alleen digitale cultuur bestuderen (kijken hoe digitalisering in alle vormen oprukt en veranderingen teweeg brengt), maar letterkundigen ook zelf in de wereld van het modellerend denken laten onderduiken. Kunnen de conceptuele analysetools waarmee wij in veel gevallen werken, worden vervangen of aangevuld door modelmatige computergestuurde analyses, en waartoe zal dat leiden? De infrastructures en middelen die we daarvoor nodig hebben (van digitale bibliotheken tot tools als Googles ngrams)¹ zijn er deels al wel, nu is het zaak dat letterkundigen er een kritische massa in gaan vormen. Met kritische massa bedoel ik dat de inrichting van zo'n minor niet als doel heeft al het letterkundig onderzoek op digitale leest te schoeien. Doel is letterkundigen op te leiden die op eigen kracht de mogelijkheden kunnen verkennen.

Want er valt nog wel het een en ander te kiezen en te bepalen. Er zijn veel meer mogelijkheden dan synthese tussen de vakgebieden zoals Valenza die voor ogen staat. In het nieuwste boek van Moretti, *The Bourgeois: Between History and Literature*, zien we nog weer een andere manier om letterkundige en informatiewetenschappelijke expertise te combineren. Moretti vermengt in dit boek *close reading* met door de computer ondersteunde vormen van *distant reading* om een cultuurhistorische analyse te maken van het fenomeen 'burgerlijkheid', en de rol die de roman in het uitdragen en verdedigen van het burgerlijk ideaal speelde.

Hoe succesvol Moretti's aanpak is, kunt je het beste zelf beoordelen. Zoals je ook zelf kunt beoordelen of de wrijving van inzichten waar het project van Van Dalen-Oskam en Bod op gebaseerd is, interessante resultaten oplevert.² En bekijk dan, als je toch een oordeel aan het vormen bent, ook even Robin Valenza's nieuwe boekproject³ om te zien waar het idee van synthese van de vakgebieden haar naartoe voert.

Literatuur

Gooding, P., Terras, M. en Warwick, C., 'The myth of the new: Mass digitization, distant reading, and the future of the book'. In: *Literary and Linguistic Computing*, 28 (2013), nr. 3:1-11.

Moretti, F. *Graphs, Maps, Trees: Abstract Models for a Literary History*, Londen 2005.

Moretti, F. *Distant Reading*, Londen 2013.

Moretti, F. *The Bourgeois: Between History and Literature*, Londen 2013.

1 URL: <<http://books.google.com/ngrams>>.

2 URL: <<http://literaryquality.huygens.knaw.nl/>>.

3 URL: <<http://www.english.wisc.edu/faculty-valenza.htm>>

- Oostendorp, M. van**, 'Geen woord te veel. Over literatuurwetenschap, naar aanleiding van *Literatuur in de wereld*', Neder-L, http://nederl.blogspot.nl/2013/10/geen-woord-te-veel.html?utm_source=twitterfeed&utm_medium=twitter, gepubliceerd 03-11-2013.
- Richardson, S.**, *Camilla, or, The History of a Young Lady*, New York 1985 (Penguin Classics).
- Valenza, R.**, 'How Literature Becomes Knowledge: A Case Study'. In: *ELH*, 76 (2009), nr. 1: 215-245.
- Zundert, J.**, 'Ongebaande paden in de stylometrie verkend', recensie van M. Kestemont, *Het gewicht van de auteur. Een onderzoek naar stylometrische auteursherkenning in de Middelnederlandse epiek*, *Textual Scholarship*, <http://www.textualscholarship.nl/?p=12134>, gepubliceerd 21-12-2012.