

PETER BOOT

## Online boekdiscussie van een afstand gelezen

*Abstract* – In the wake of Franco Moretti’s notion of distant reading, much has been written about the potential for a quantitative and computational turn in the study of literary phenomena. This approach seems especially suitable to studying literary reception, among other reasons because the number of reception documents has risen sharply with the advent of Web 2.0. In this article I sketch the landscape of online book discussion, and discuss the research potential of a large and representative corpus of online responses to books. I discuss some of the implications of the quantitative-computational turn for literary studies.

### 1 Inleiding

De ideeën van Franco Moretti over *distant reading* (Moretti 2005) hebben bij veel letterkundig onderzoekers weerklank gevonden. Moretti stelt dat in het onderzoek van culturele verschijnselen met een zekere schaalgrootte lezen alleen niet meer toereikend is en vraagt om een methodiek waarbij concentratie op bepaalde, meetbare aspecten verschijnselen zichtbaar maakt die ook voor de geduldige lezer onzichtbaar blijven. Eén van de redenen waarom Moretti’s ideeën aanslaan, is de overvloed aan digitaal materiaal die dankzij de digitaliseringsinspanningen van Google, de DNB en anderen de laatste tien jaar beschikbaar is gekomen. De wanverhouding tussen de hoeveelheid beschikbaar en de hoeveelheid daadwerkelijk bestudeerd bronmateriaal springt pijnlijk in het oog. ‘What do you do with a million books?’ was naar aanleiding hiervan de vraag van Gregory Crane (2006) en het antwoord van Clement e.a. was ‘How not to read a million books’ (2008): omdat niemand een miljoen boeken kan lezen, is de inzet van de computer de enige manier om die grote hoeveelheden digitaal materiaal te kunnen analyseren. Maar digitalisering maakt hier alleen een problematiek zichtbaar die ook voorheen al speelde, en die de context leverde voor Moretti’s oorspronkelijke formulering van het *distant reading*-begrip: de hoeveelheid literatuur die buiten het blikveld van historisch en vergelijkend letterkundigen blijft doordat die zich noodgedwongen beperken tot een kleine gecanoniseerde selectie van werken (Moretti 2013a). Hetzelfde geldt voor de analyse van (online) digitale cultuuruitingen: de hoeveelheid relevant digitaal materiaal op Twitter, weblogs, en andere sites maakt computationele analyse noodzakelijk.<sup>1</sup>

Natuurlijk is Moretti niet de eerste die in de studie van literaire verschijnselen een kwantitatieve benadering voorstaat. In de neerlandistiek is deze benadering eerder voorgesteld in de invloedrijke school van Van Rees en Verdaasdonk. Vaak ging daarbij de aandacht voornamelijk uit naar niet-tekstuele, sociale aspecten.

<sup>1</sup> Prescott betoogt welsprekend dat de Digital Humanities zich niet moeten beperken tot de studie van gecanoniseerde culturele objecten maar juist een belangrijke taak hebben in de studie van eigentijdse digitale cultuur (Prescott 2012).

ten, zoals de kans voor een auteur om de P.C. Hoofd-prijs te winnen op basis van eerder gewonnen prijzen, jurylidmaatschappen en leeftijd (Verdaasdonk 2008b). Maar Verdaasdonk besteedde ook aandacht aan tekstuele verschijnselen, zoals aantal en soort citaten in recensies als functie van de reputatie van de auteur (Verdaasdonk 2008c) of verwijzingen in recensies naar andere auteurs (Verdaasdonk 2008a).

Verdaasdonk c.s. concentreerden zich in hun onderzoek op de kritiek, in overeenstemming met de centrale these van de literatuursociologie dat literaire reputatie niet is gebaseerd op het literaire werk zelf maar wordt gemaakt door critici en andere institutionele actoren. Misschien is het ook zo dat kritiek, een proces waarbij in een sociale context meerdere personen reageren op een werk maar vooral ook op elkaar (Van Rees 1985), een inherent meer sociologisch verschijnsel is dan de literaire tekst, en zich daarom eerder leent voor kwantitatieve bestudering. Een andere reden kan zijn dat hulpmiddelen voor de automatische analyse van tekst pas de laatste jaren ter beschikking beginnen te komen. Verdaasdonk telde met de hand de verschijnselen waarin hij geïnteresseerd was. In die omstandigheden kan het een onmogelijkheid lijken om verifieerbare uitspraken over omvangrijkere literaire teksten te doen. Nu computationele middelen voor literaire tekstanalyse niet meer ondenkbaar zijn, is er misschien minder reden dan in de tijd van Verdaasdonk om de literaire tekst zelf te beschouwen als een *black box* waarover geen objectieve uitspraken kunnen worden gedaan.

In deze bijdrage wil ik kijken naar mogelijkheden van een morettiëske aanpak in de studie van de letterkunde, en meer in het bijzonder het receptieonderzoek. Als kenmerkend voor deze aanpak beschouw ik de concentratie op een beperkt aantal meetbare of objectiveerbare aspecten van teksten, die met behulp van onder andere taalkundige analyse, statistiek en visualisatietechnieken ‘van een afstand’ worden gelezen. Daardoor zouden patronen zichtbaar moeten worden die voor de ‘van dichtbij’ lezende onderzoeker verborgen blijven. Als voorbeelden van dit soort technieken kunnen gelden de grafieken, de projectie op geografische kaarten en de boomstructuren van *Graphs, Maps and Trees* (Moretti 2005) of sociaal-netwerkdigrammen (Moretti 2013b). De casus waar ik me op richt is onderzoek naar de (Nederlandse) online-boekdiscussie. Online-discussie over boeken is een verschijnsel met een zodanige omvang dat door computationele hulpmiddelen ondersteunde kwantitatieve benadering noodzakelijk lijkt; het is een verschijnsel ‘te groot voor het blote oog’ (Boot 2008). Ik schets het terrein van de online-boekdiscussie en geef aan hoe de bouw van een corpus van online-boekrespons aan zulk onderzoek een bijdrage zou kunnen leveren. Vervolgens geef ik een aantal voorbeelden van analyses die daardoor mogelijk worden. Ik sluit af met een aantal overwegingen over de implicaties van de voorgestelde vorm van onderzoek voor de methoden van de neerlandistiek en de geesteswetenschappen meer algemeen.

## 2 Online boekrespons

Het landschap van de online-boekdiscussie is gevarieerd en in beweging. Het is niet gemakkelijk om een consistente typologie van sites te ontwerpen, onder andere omdat sites en sitetypes in constante ontwikkeling verkeren, of omdat sites

verschillende soorten functionaliteit kunnen bieden (bijvoorbeeld discussie en boekbespreking) die zowel gecombineerd als afzonderlijk kunnen voorkomen (zie Boot 2011 voor een bespreking van de problematiek). Het terrein is veel wijder dan de literaire kritiek die vaak het uitgangspunt vormt voor beschouwingen over online-boekdiscussie (Joosten 2010; McDonald 2007). In de inventarisatie die ik hier geef, betrek ik in principe alle online-respons op boeken, los van de vorm die deze krijgt, het medium dat er voor gebruikt wordt, het geïntendeerde publiek, de kwalificaties van de respondent of het oogmerk van de respons. Alle boekgenres zijn in principe interessant, ook populaire literatuur en non-fictie, al zal ik reacties op praktische teksten als bijvoorbeeld kookboeken en schoolboeken buiten beschouwing laten. Ik ben in dit verband vooral geïnteresseerd in respons van lezers, en minder in het gebruik van internet als middel voor auteurs om zichzelf te positioneren. Maar er is geen harde grens tussen beide rollen, en veel auteurs zijn ook actief als online-criticus (zie voor internetactiviteiten van auteurs Dera 2012).

Er zijn verschillende redenen waarom deze sites aandacht verdienen. De sites tonen veranderingen in processen van canonisering en repertoirevorming in het algemeen die belangrijke gevolgen kunnen hebben voor letterkunde en cultuur.<sup>2</sup> Respons op literatuur leert ons bovendien wat literatuur voor lezers kan betekenen. In het bijzonder toont respons hoe de besprekers literatuur kunnen gebruiken om zichzelf te presenteren en positioneren.<sup>3</sup> Daarnaast documenteren boekdiscussiesites receptie op een schaal die voorheen ondenkbaar was. Het is daarom misschien verrassend dat de meeste van de sites die ik hieronder zal noemen vanuit de neerlandistiek nog geen enkele aandacht hebben gekregen.

### *Boekgebaseerde sociaal-netwerksites*

Boekgebaseerde sociaal-netwerksites zijn sites waar de leden reacties plaatsen op de boeken die ze hebben gelezen en op basis van gedeelde voorkeuren vriendschappen kunnen aangaan. Voorbeelden in het buitenland zijn het onlangs door Amazon overgenomen Goodreads en in Duitsland lovelybooks.<sup>4</sup> In Nederland zijn de bekendste sites watleesjij.nu en dizzie.nl. Als voorbeeld bespreek ik kort watleesjij.nu, op basis van een eerdere beschrijving in Boot 2012b.

Op watleesjij.nu zijn ongeveer 9000 lezers actief, die ongeveer 30.000 reacties op boeken hebben geplaatst. De meest besproken boeken zijn populaire recente (vertaalde) literaire fictie; daarnaast worden veel thrillers besproken. Motief voor bezoek aan de site is veelal de wens om nieuwe boeken te leren kennen: 'ik baal altijd als ik een boek lees dat mij zwaar tegenvalt. Dat vind ik zo zonde van mijn tijd. Daarom: Wat een handige site is dit!' De reacties variëren van uiterst beknopte uitroepen tot langere besprekingen. Onder de frequente bezoekers van de site worden de beknopte commentaren door eenmalige bezoekers niet altijd op prijs gesteld: 'jammer dat er zoveel onbekende eendagsvliegers met een regelige recensies zijn'. Bezoekers kunnen vriendschappen sluiten met andere bezoekers, maar

<sup>2</sup> Zie voor het begrip repertoirevorming Andringa 2006.

<sup>3</sup> Zoals terecht opgemerkt door een anonieme reviewer van dit artikel.

<sup>4</sup> <http://www.goodreads.com>, <http://www.lovelybooks.de>.

dat is een faciliteit waar lang niet iedereen gebruik van maakt. Er zijn mensen met tientallen vriendschappen, maar ook actieve gebruikers die op de site geen vriendschappen aangaan. Vrienden kunnen elkaar berichtjes sturen, en sommige leden maken enthousiast van die mogelijkheid gebruik.

In het wetenschappelijk onderzoek is aan boekgebaseerde sociaal-netwerksites vooral aandacht besteed in verband met de mogelijkheden die ze bieden voor geautomatiseerde boek-aanbevelingen (Koolen e.a. 2012). Verboord (2011) kijkt naar de invloed van Goodreads op de boekenmarkt. Ook is er over deze sites wel geschreven vanuit het perspectief van de bibliotheek en van leesbevordering.

### *Forums*

Lang voordat er sociaal-netwerksites bestonden werd al over literatuur gediscussieerd op forumsites. Soms gebeurt dat op aan literatuur gewijde deelforums van grote sites, zoals het TTK (Theater, taal en kunst) subforum van fok.nl,<sup>5</sup> soms op volledig aan boeken gewijde forums. Het grootste Nederlandstalige boekenforum is Ezzullia.<sup>6</sup> Op de site staan bijna een half miljoen berichten, afkomstig van ruim 3000 leden. De site is georganiseerd in verschillende deelforums, sommige bedoeld voor algemene boek-gerelateerde conversatie (wat ben je nu aan het lezen?), andere voor boekadviezen, sommige voor acties waarbij gezamenlijk bepaalde boeken worden gelezen, en weer andere voor discussie over specifieke auteurs. Voor bijdragen met meer pretentie bestaat een afzonderlijk subforum voor recensies. Op Ezzullia worden alle genres besproken, hoewel het zwaartepunt lijkt te liggen bij de thrillers.

Meer nog dan de sociaal-netwerksites zijn forums voor een belangrijk deel gezelligheidsplatforms, waar mensen vaak eerder aan het gesprek meedoen om het contact dan om iets over een boek te zeggen. Niet zelden leidt dat tot discussiebijdragen als 'ik ben nu op bladzij 80', 'ik ga het boek lezen zodra ik boek x uit heb'. Behalve een boekenforum is Ezzullia dan ook een plaats waar een groep bekenden elkaar treft rond het thema boeken.

Naast het forum omvat Ezzullia overigens ook een website waar door een redactie van vrijwilligers over boeken wordt geschreven. Hier verschijnen interviews, recensies, boektips, nieuwsberichten en columns. Dit is een verschijnsel dat bij veel meer boekensites optreedt: vaak is er een forum waar iedereen kan meepraten, en een ander deel van de site waarvan de inhoud door een redactie wordt beheerd. Vanuit de sociologie van de literatuurbeschuwing is dat een interessant verschijnsel: het internet heeft het mogelijk gemaakt dat niet-professionals publiceren over boeken, maar ook onder niet-professionals zijn er kennelijk gradaties, en corresponderende verschillen in status.

### *Boekhandels*

Boekbesprekingen in online-boekhandels zijn waarschijnlijk de bekendste vorm van online-boekdiscussie. Vooral over de besprekingen bij Amazon is veel ge-

<sup>5</sup> <http://forum.fok.nl/forum/15>.

<sup>6</sup> <http://www.ezzullia.nl/forum/index.php>.

schreven, met name vanuit een marketingperspectief. Boekbesprekingen zijn, net als besprekingen van andere producten, veel bestudeerd met het oogmerk automatisch hun strekking te bepalen (De Smedt en Daelemans 2012b). Andere aspecten die veel aandacht hebben gekregen zijn de betrouwbaarheid van de besprekingen: hoe kan de lezer weten dat een positieve recensie niet afkomstig is van de auteur van het boek, of een negatieve recensie van één van diens vijanden?

In het Nederlands taalgebied is waarschijnlijk bol.com de boekhandel waar de meeste boekbesprekingen van klanten worden geplaatst. Voor een onderzoekje naar de receptie van Arnon Grunberg vond ik in december 2011 271 reacties op zijn boeken. Die aantallen reacties maken bol.com tot een belangrijke bron voor meningen van lezers over literatuur. Maar we zouden, net als bij andere sites overigens, veel meer willen weten over de personen die achter deze meningen schuil gaan, en de mate van overlap met andere sites. De mogelijkheden daartoe zijn echter heel beperkt.

### *Weblogs*

Het terrein van de weblogs valt moeilijk precies af te grenzen van de recensiesites en online-literatuurmagazines. Een mogelijk criterium om de weblogs te onderscheiden van andere sites is de afwezigheid van institutionalisering: een prototypische weblog krijgt geen subsidie, heeft geen redactie of raad van advies, en geeft zelf geen boeken uit. Een grensgeval wordt gevormd door de weblogs die weliswaar zelf geen instituut vertegenwoordigen, maar waarvan de auteurs wel professioneel of semiprofessioneel een rol spelen in het literaire veld. Sommige critici verzamelen bijvoorbeeld hun recensies op hun weblog, soms vergezeld van andere overwegingen of berichten over de boekenwereld.

Een boekenweblog is in zijn meest zuivere vorm een website waar één lezer, soms een groepje lezers, verslag doet van zijn leeservaringen. Maar juist omdat een weblog een persoonlijke onderneming is, zal de inhoud vaak gestuurd worden door de combinatie van belangstellingen van de auteur. Zo bestaan er weblogs over boeken en muziek, boeken en reizen, of boeken en koken. Bij het persoonlijke karakter hoort ook dat de besproken boeken en de aard van de besprekingen afhankelijk zijn van de persoon van de blogger.<sup>7</sup> De gemiddelde weblogger wordt minder gestuurd door het actuele boekenaanbod en de hypes in boekenland dan de lezers van watleesjij.nu.

### *Recensiesites en online-magazines*

Kenmerkend voor deze sites, anders dan de tot nog toe genoemde sites, is dat ze zichzelf zien als onderdeel van de letterkundige wereld en ambiëren daarin een rol te spelen. Ik zal niet proberen een precies onderscheid te maken tussen recensiesites en online-magazines. Het spectrum loopt van een zuivere recensiesite als Recensieweb tot Ooteote (dat zich beschrijft als 'online informatieknoop-

<sup>7</sup> Ik noem een paar voorbeelden uit een onafzienbare collectie: 'las!' van Marc van Oostendorp, Boeklog van Ijsbrand van den Berg, Anna's leesreis van Anna van Gelderen, en de veel genoemde Achille van den Branden (<http://marc-las.blogspot.nl/>, <http://boeklog.info/>, <http://annavangelderden.blogspot.nl/>, <http://www.achillevandenbranden.net/>).

punt').<sup>8</sup> Ik reken tot deze categorie ook De Reactor ('platform voor literaire kritiek'), Literair Nederland ('website voor lezers en leesclubs'), de contrabas ('weblog over literatuur'), en Tzum ('literaire weblog'). De verhouding tussen opinies en berichtgeving ligt op elk van deze sites anders. Niet alle sites richten zich alleen op boeken: De Recensent, 8Weekly en het Vlaamse Cutting Edge bespreken ook films, muziek en andere media. Een bespreking van ambities en mogelijkheden van deze sites valt buiten het kader van dit artikel.<sup>9</sup>

### *Genrespecifieke sites*

Veel sites die zich richten op boeken zijn specifiek geïnteresseerd in bepaalde genres. Er bestaan sites die zich speciaal richten op thrillers,<sup>10</sup> op *fantasy* (al dan niet in combinatie met *science fiction*)<sup>11</sup> of op romantische boeken.<sup>12</sup> Veel van die sites omvatten een forum, een onderdeel voor recensies, en vaak ook interviews met auteurs, nieuwsberichten, etc. Vaak is een site niet alleen een informatiebron, maar ook een plaats waar gelijkgezinden elkaar ontmoeten. Er zijn ook sites die zich vooral richten op boeken over een specifiek onderwerp, zoals de tweede wereldoorlog of religie.<sup>13</sup>

### *Twitter*

Twitter is in deze context in de eerste plaats van belang omdat veel sites uit de overige categorieën Twitter gebruiken om geplaatste artikelen en recensies aan te kondigen. Anderen gebruiken Twitter om berichten van verschillende sites te verzamelen.<sup>14</sup> Het bekendste zelfstandige gebruik van Twitter voor boekdiscussie is de #TwitLit-leesclub, gemodereerd door Bas Heijne. In deze leesclub is de afgelopen jaren een aantal boeken besproken, vaak inclusief een sessie waarin de auteur van het besproken werk aan de discussie deelnam.<sup>15</sup> Twitter wordt daarnaast gebruikt om een werk of auteur onder de aandacht te brengen door fragmenten uit het werk te tweeten, eerder door liefhebbers dan door de auteur zelf.<sup>16</sup> En tenslotte kan elke twitteraar, buiten alle structuren om, een tweet de wereld insturen over het boek of de auteur die hij toevallig leest. In dat geval zijn die auteurs in het voordeel die regelmatig publiceren (bijvoorbeeld Grunberg met zijn dagelijkse Voetnoot voorop de Volkskrant).

8 <http://recensieweb.nl>, <http://ooteoote.nl>.

9 Zie Absillis (2010) voor een bespreking van een aantal van deze sites. Boot (2012a) bespreekt 8Weekly.

10 <http://inejacet.nl>, <http://crimezone.nl>

11 <http://magictales.nl>, <http://www.ff-leesclub.nl>, <http://www.purefantasy.nl>

12 <http://www.romantischeboeken.nl>, <http://www.chasingdreams.nl>

13 <http://www.go2war2.nl/Recensies/>, <http://www.chroom.net/>

14 <https://twitter.com/boekrecensie>.

15 <https://twitter.com/#!/search/%23TwitLit>.

16 Bijvoorbeeld <https://twitter.com/FritsVanEgters>, voor een enkel werk, of <https://twitter.com/twitcit> voor willekeurige werken.

### *Diversen*

De hier genoemde sitetypen zijn de belangrijkste typen waar het gaat om boekdiscussie. Maar er zijn nog andere soorten sites waar meningen over literatuur worden gegeven die ook potentieel bijdragen aan de repertoirevorming. Een belangrijke klasse vormen de formele of minder formele nieuwssites, zoals Boekendingen of de (verdwenen) Papieren man.<sup>17</sup> Ook andere sites met een informatief oogmerk dragen bij aan de meningsvorming. Wikipedia is een belangrijk voorbeeld, maar infonu.nl is een andere site met de pretenties van een naslagwerk.<sup>18</sup> Evenzeer van belang zijn sites met boekverslagen van scholieren en studenten. Nog weer een heel andere manier van reageren op literatuur, die ik hier verder buiten beschouwing zal laten, is creatieve verwerking in de vorm van *fan fiction*.<sup>19</sup> Ook online-heruitgave van letterkundig werk kan gezien worden als een statement over de waardering die een werk toekomt. Wanneer op de site De Leestrommel Indische damesromans worden gepubliceerd, of op 'De open ruimte' het volledig werk van Rudy Cornets de Groot, geven de samenstellers daarmee aan dat het gaat om cultuurgoed dat ten onrechte dreigt te worden vergeten.<sup>20</sup> Tenslotte zijn ook 'vlogs' (video blogs) een manier om aan letterkundige discussie deel te nemen. Een site waar dat op structurele wijze gebeurt, is Why I love this book,<sup>21</sup> waar mensen in één minuut uiteen moeten zetten wat ze goed vinden aan een boek, maar er zijn ook veel individuen die boekbesprekingen in videoformaat op YouTube zetten.<sup>22</sup>

### *Boekdiscussiesites als netwerk*

Een belangrijk aspect van de wereld van de online-boekdiscussie is dat de meeste deelnemers actief zijn op meerdere sites. De boekbesprekers van 8Weekly bijvoorbeeld hebben vaak ook een eigen weblog, of ze schrijven ook recensies voor andere sites. Bij watleesjij.nu hebben lezers vaak ook een account op andere boekensites, waar ze soms ook dezelfde besprekingen plaatsen. Soms sluiten ze er zelfs vriendschap met dezelfde mensen die ze kennen van watleesjij.nu (Boot 2012b). Hetzelfde geldt voor de leden van forum Ezzullia. Gedeeltelijk heeft dit zonder twijfel te maken met het uitproberen van de verschillende platforms, maar vaak worden ook meerdere platforms gedurende langere tijd naast elkaar gebruikt.

## 3 Een corpus van boekrespons

Zoals boven al aangegeven kunnen boekdiscussiesites een belangrijke bron zijn voor neerlandistisch en literatuurwetenschappelijk onderzoek. Omdat op de diverse typen sites op heel verschillende manieren over boeken wordt geschreven, is het van belang om in dat onderzoek zoveel mogelijk verschillende sites en soorten

17 <http://www.boekendingen.nl/>, <http://www.depapierenman.be/>.

18 <http://recensies.infonu.nl/kunst-en-cultuur/>.

19 De belangrijkste Nederlandse site is <http://www.fanfic.nl/>.

20 <http://www.leestrommel.nl/>, <http://www.cornetsdegroot.com/>

21 <http://www.whyllovethisbook.com/>.

22 Bijvoorbeeld <http://www.youtube.com/user/booksandquills>.



sites te betrekken. Voor een onderzoeksaanpak die computationele hulpmiddelen wil gebruiken, zijn bovendien van elke site veel teksten nodig. Dat impliceert dat er alle reden is om een corpus bijeen te brengen van uit zoveel mogelijk bronnen afkomstig en divers materiaal dat ons in staat stelt om empirisch onderzoek te doen naar de receptie van literatuur in het tijdperk van het Internet.

TABEL 1 Samenstelling testcorpus

<i>Bron</i>	<i>Artikelgenre</i>	<i>Aantal reacties</i>	<i>Aantal woorden</i> (*1000)
8weekly magazine	recensie	2273	1512
weblogs	blog post	6952	3578
watleesjij.nu	lezersrecensie	28037	2515
Crimezone boek beschrijving	beschrijving	3698	462
Crimezone recensies	recensie	3696	1622
Crimezone recensies door bezoekers	lezersrecensie	9163	1537
Grunberg recensies (gedrukt)	recensie	196	187
NRC webshop beschrijving	beschrijving	1345	198
NRC recensies (gedrukt)	recensie	1226	1133
<b>Totaal</b>		<b>56586</b>	<b>12744</b>

Om de mogelijkheden van zo'n collectie nader te bekijken heb ik een testcorpus bijeengebracht waarvan de samentelling wordt beschreven in Tabel 1. Het corpus omvat nu ongeveer 56000 reacties op boeken bestaande uit in totaal bijna 13 miljoen woorden. De recensies van weblogs zijn afkomstig van acht verschillende, merendeels tamelijk intellectuele weblogs.<sup>23</sup> Crimezone is een site over thrillers waarvandaan ik verschillende tekstsoorten heb gedownload: boekbeschrijvingen, waarschijnlijk afkomstig van uitgevers, in opdracht geschreven recensies en gebruikersrecensies. Ter vergelijking bevat het corpus uit de gedrukte pers een verzameling van recensies van werken van Grunberg, gedownload van Grunbergs site, en twee soorten teksten uit de boekwinkel van de NRC: boekbeschrijvingen en een verzameling uit de krant overgenomen recensies. Ik zou dit corpus nog willen uitbreiden met meer soorten sites (bijvoorbeeld forums, Twitter, sites met boekverslagen), meer soorten tekst (nieuwsberichten over boeken, encyclopedische teksten), en meer boekgenres (bijvoorbeeld *fantasy*, romantische boeken, sf, poëzie). Maar ook met de huidige samenstelling valt al een aantal mogelijkheden te demonstreren.

De gegevens zijn verzameld in de periode van oktober 2012 tot januari 2013, met uitzondering van de gedrukte besprekingen van Grunberg, die werden gedownload in februari 2012. De besprekingen zelf zijn in veel gevallen ouder en gaan bijvoorbeeld in het geval van 8Weekly een enkele keer terug tot 2000. Ten behoeve van de gegevensverzameling werden site-specifieke scripts geschreven. In sommige gevallen werd daarvoor eerst de hele site gedownload. De scripts analyseerden de pagina's van de sites en extraeerden daaruit de tekst van de bespre-

23 Blijkens de besproken boeken en de aard van de discussies.



king en metadata, voor zover beschikbaar (ISBN, datum, beoordeling, auteur review, etc.).<sup>24</sup>

Omdat boekbesprekingen een belangrijk deel van hun betekenis ontlenden aan context, zou een corpus meer moeten zijn dan een verzameling van enkel teksten. De metadata zouden in elk geval gegevens moeten bevatten over de site en de deelcollectie op de site waaruit een bericht afkomstig is, en over het bericht waar een bericht op reageert (als dat er is). Ook van belang zijn gegevens over het boek waar het bericht op reageert (inclusief bijvoorbeeld genre, zodat selectie daarop mogelijk wordt), en gegevens over de bespreker, inclusief, als die beschikbaar zijn, gegevens over favoriete boeken en over zijn of haar op boekensites aangegane vriendschappen. Met die laatste gegevens zouden we bijvoorbeeld kunnen onderzoeken in hoeverre vriendschap gebaseerd is op overeenkomsten in leesgedrag of in waardering van boeken (of misschien juist zulke overeenkomsten tot gevolg heeft). Of misschien waarderen mensen die vriendschappen sluiten wel andere boeken dan mensen die dat niet doen. Overigens zouden privacy-overwegingen het verzamelen van deze gegevens nog wel problematisch kunnen maken.

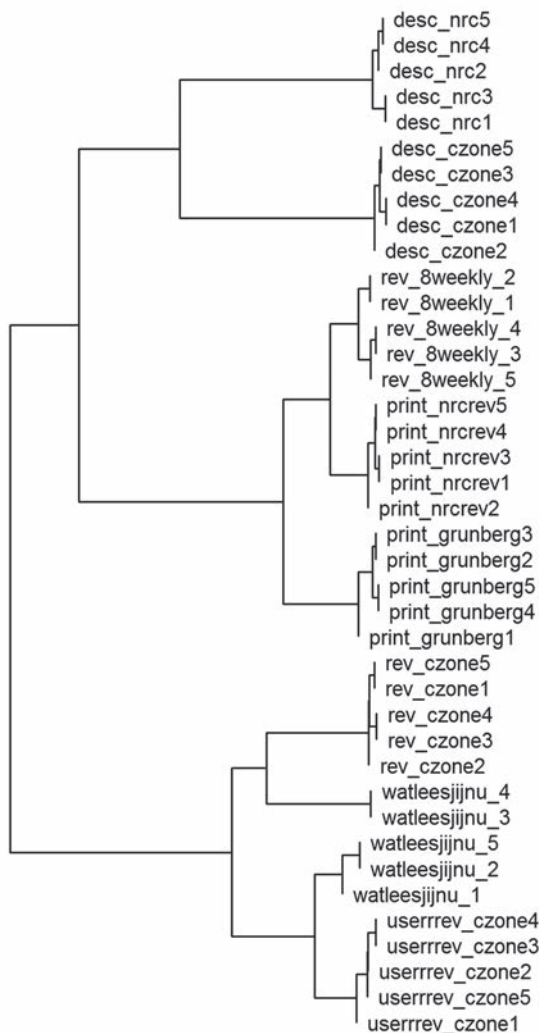
Een corpus als dit kan gebruikt worden voor het onderzoeken van heel diverse vragen: voor toegepaste vragen als de receptie van een enkel werk of auteur, maar ook voor meer theoretische vragen als de aard van de literatuuropvattingen die op de verschillende platforms worden gehanteerd of bijvoorbeeld de rol van sociale status bij het prijzen of verwerpen van bepaalde werken of genres. De laatste vraag is op het moment extra interessant omdat we in een situatie verkeren waarin andere dan de traditionele stemmen (critici, leraren, academici, mediapersoonlijkheden) de mogelijkheid krijgen om mee te praten over de beoordeling van literatuur.

Maar wellicht belangrijker dan het direct ondersteunen van dit soort inhoudelijk onderzoek is dat met de totstandkoming van een omvangrijk corpus van receptieteksten de mogelijkheid ontstaat om patronen zichtbaar te maken. Wat we met een corpus kunnen doen is het ontwikkelen en valideren van tools die ons in staat stellen die patronen op te sporen, bijvoorbeeld op het vlak van de gebruikte argumentatietechnieken, de normen die voor geslaagde literatuur worden aangelegd, of de media en auteurs die worden beschouwd als autoriteiten. Dat soort patronen zou je vervolgens in verband willen brengen met bijvoorbeeld voorkeuren voor bepaalde boeken.

Om de mogelijkheden voor computationele analyse op een corpus als dit te illustreren laat ik een aantal analyses zien die ik op het testcorpus heb uitgevoerd.

<sup>24</sup> De scripts moeten voor elke site opnieuw worden ontwikkeld, omdat de functionaliteit van de diverse sites verschilt, omdat verschillende softwarepakketten worden gebruikt voor het beheer van de sites (en in het geval van de weblogs verschillende templates), en omdat elke site eigen conventies hanteert met betrekking tot de inrichting van de site en van de pagina's. In een enkel geval waren zgn. rss-feeds (rss: Really Simple Syndication, of ook wel: RDF Site Summary) beschikbaar, dat wil zeggen een weergave van de gegevens zonder opmaak, waardoor het proces van extractie eenvoudiger wordt. Het resultaat van de extractie is een verzameling tekstbestanden voor de besprekingen en een verzameling csv-bestanden (comma-separated values) voor de andere gegevens. Deze laatste bestanden kunnen worden bekeken in spreadsheetprogramma's zoals Excel of nader worden geanalyseerd in andere programmatuur. De collectie gegevens zoals ik die nu heb verzameld is een testcollectie en berust voor een deel op ad-hoc beslissingen. Het resultaat is in sommige opzichten minder gestructureerd dan gewenst. Ik deel de gegevens echter graag met collega-onderzoekers die in de mogelijkheden geïnteresseerd zijn.

Fig. 1 Clustering van responsteksten gebaseerd op 300 meest frequente woorden.



### Stilistische clustering van site typen

Kunnen computationele tools gebruikt worden om de teksten afkomstig van verschillende sitetypen te onderscheiden? Het antwoord is, met enig voorbehoud, ja. Fig. 1 toont het resultaat van een clusteranalyse, uitgevoerd met de stylometrische tools voor R, ontwikkeld door Maciej Eder & Jan Rybicki (2011). De clusteranalyse maakt een zo goed mogelijke groepering van bestanden op basis van overeenkomst, en is uitgevoerd op vijf bestanden van elk 20000 woorden uit elke collectie. De analyse is, zoals dat in modern stylometrisch onderzoek gebruikelijk is, gebaseerd op het gebruik van de meest frequente woorden (dus vooral functiewoorden: persoonlijk voornaamwoorden, voorzetsels, voegwoorden, etc.) We zien hier hoe de software in staat is tekstgroepen uit dezelfde collecties correct te groe-

peren: de teksten van elke collectie, met uitzondering van die van *watleesjij.nu*, zijn kennelijk stilistisch tamelijk consistent. Opmerkelijker is dat de software er ook tot op redelijke hoogte in slaagt teksten succesvol te groeperen op basis van hun tekstgenre: we zien bovenin de beschrijvende teksten van de Crimezone site en de NRC-webshop, daaronder de traditionele recensies, zowel gedrukt (Grunberg en NRC) en online (8weekly), en onderaan de besprekingen van het Crimezone-platform en die van *watleesjij.nu*. Een interessante vraag voor nader onderzoek is waarom die bij elkaar in de buurt terecht komen. Wat ermee te maken kan hebben is dat er een zekere personele overlap bestaat tussen de reviewers op beide platforms. Waar het zeker niet mee te maken heeft is met de inhoud van de besprekingen, ook al worden op *watleesjij.nu* veel thrillers besproken: als dezelfde clustering wordt uitgevoerd op basis van alleen de woordsoorten, waarbij dus de factor inhoud is uitgeschakeld, is het resultaat vrijwel identiek.<sup>25</sup>

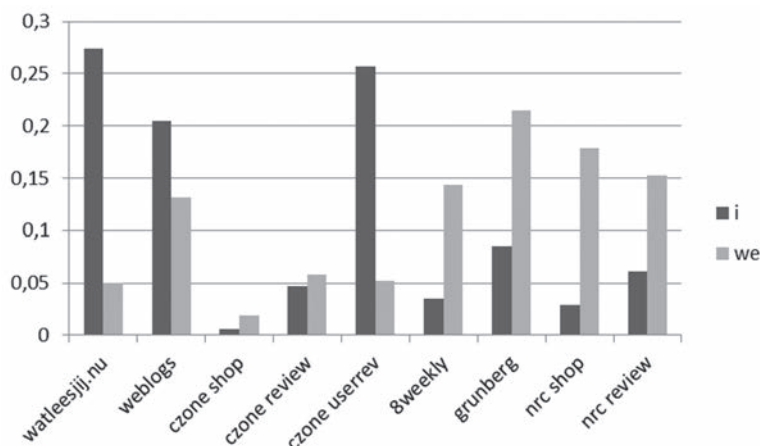
Een clusterdiagram kan dus, tot op zekere hoogte, worden gebruikt om sites te groeperen op genre. Meer fundamenteel roept het diagram de vraag op hoe het komt dat verschillen in frequent gebruikte woorden, en zelfs woordsoorten, gebruikt kunnen worden om sites en genres te onderscheiden. Wat is het dat maakt dat de NRC-site andere woordsoorten gebruikt voor de aanbeveling van boeken dan CrimeZone, maar dat ze toch meer op elkaar lijken dan op recensieteksten? En om welke woordsoorten gaat het dan?

### *Frequentieverschillen in woordcategorieën*

De hierboven getoonde clustering is lastig te interpreteren in termen van direct zichtbare eigenschappen van de teksten. Daarvoor kunnen we meer eenvoudige hulpmiddelen gebruiken, bijvoorbeeld de tellingen van de LIWC (Linguistic Inquiry and Word Count; Zijlstra e.a. 2004), een hulpmiddel ontwikkeld door psychologen dat woorden telt in bepaalde syntactische en semantische categorieën. Ik laat twee voorbeelden zien. Grafiek 1 toont het relatieve gebruik van eerste persoon enkelvoud en meervoud in de verschillende collecties. Heel duidelijk zichtbaar is dat op de userplatforms, *watleesjij.nu* en gebruikersbesprekingen op Crimezone de eerste persoon enkelvoud domineert: mensen spreken daar voor zichzelf. In iets mindere mate geldt dat ook voor de weblogs. In de kritiek, zowel in de besprekingen in de NRC als in de online kritiek van 8Weekly, overweegt het 'we'. Een verklaring zou gezocht kunnen worden in de mate waarin besprekers zich autoriteit aanmeten. Een andere mogelijkheid is dat de ik-gebruikende lezers hun lectuur eerder op zichzelf betrekken dan dat ze zoeken naar meer algemene betekenissen. Om die hypothesen nader te onderzoeken is toegang tot de besprekingen zelf noodzakelijk. Overigens slagen, opmerkelijk genoeg, de beschrijvingen in de Crimezone-winkel erin om de eerste persoon vrijwel geheel te vermijden. Redactioneel beleid speelt hier waarschijnlijk een belangrijke rol.

25 Om precies te zijn: bi-grams van woordsoorten, waarbij dus gekeken wordt naar de woordsoorten van opeenvolgende woorden. Hierbij zijn de woordsoorten bepaald met behulp van de *Pattern* tools (Desmedt en Daelemans 2012a). Teksten waarbij de woorden waren vervangen door een aanduiding van hun woordsoorten werden vervolgens aangeboden aan de stilometrische toolset van Eder en Rybicki. Deze toolset biedt de mogelijkheid om de analyses uit te voeren op n-grams (groepen van n woorden) in plaats van op losse woorden.

Grafiek 1 Genormaliseerde frequenties eerste persoon enkelvoud en meervoud.



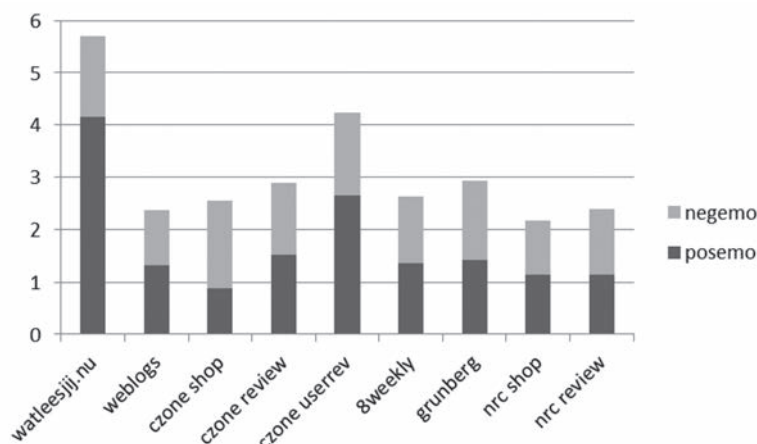
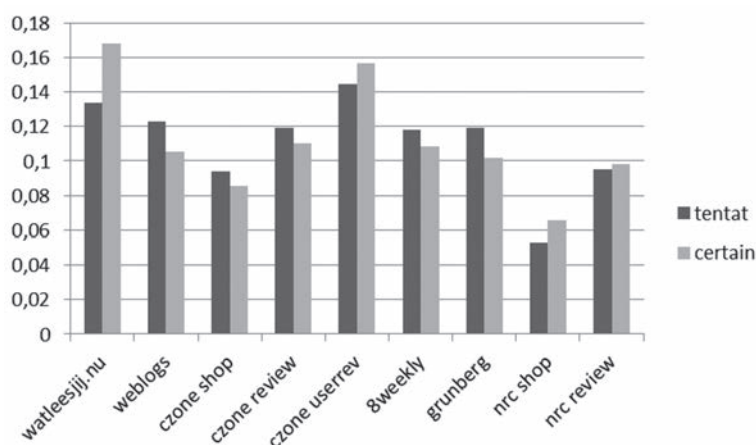
Ook het gebruik van emotiewoorden vertoont herkenbare patronen. Grafiek 2 toont de frequenties van positieve en negatieve emotiewoorden in de verschillende collecties, gebaseerd op de LICW-categorieën. Watleesjij.nu en de Crimezone-userreviews springen er weer uit in termen van emoties, maar vooral positief: het zijn sites waar mensen eerder juichen over een mooi boek dan waar een afgewogen oordeel geveld wordt. In deze grafiek horen de weblogs niet bij de ‘informele’ sites, misschien omdat de weblogs in het testcorpus het meer bedachtzame deel van het genre vertegenwoordigen. Interessant is ook de hoeveelheid negatieve emotiewoorden in de Crimezone-boekbeschrijvingen. Misschien is die het resultaat van het feit dat woorden als ‘angst’ en ‘spanning’ voor thrillers kunnen fungeren als aanbevelingen, en ook in beschrijvende zin kunnen worden gebruikt.

Metingen zoals deze leveren dus nooit eenduidige antwoorden. Ze maken verschijnselen zichtbaar die vragen om een verklaring. Zo’n verklaring kan worden gezocht in de aard van de sites, maar het is ook mogelijk dat het meetinstrument tekort schiet (bijvoorbeeld doordat de LICW bepaalde emotiewoorden zou missen, of dat het tellen van individuele woorden voor het bepalen van de mate van emotionaliteit niet de geëigende methode is).

### *Verwijzingen naar oeuvre en stijl*

Ook wenselijk is een automatische analyse van aspecten die we in boekbesprekingen verwachten, zoals verwijzingen naar eerder werk van de auteur of naar stilistische aspecten. Ik heb een aantal zeer experimentele tellingen uitgevoerd op basis van lijsten van relevante woorden en woordgroepen.<sup>26</sup> Het resultaat wordt getoond in Grafiek 3. We zien veel aandacht voor stijl in de kritieken (NRC, Grunberg, 8Weekly, ook in de Crimezone-reviews), veel aandacht voor oeuvre in de besprekingen van werk van Grunberg. Maar dit is niet meer dan een allereerste

<sup>26</sup> Voor stijl: de woorden stijl en stilistisch, een aantal termen als woordkeus, zinsbouw, taalgebruik, en namen van stijlfiguren. Voor oeuvre: formuleringen als eerste/tweede/vorige/laatste boek/roman, opnieuw een [woord] boek, en dergelijke. Dat dit toepasselijke zoekuitdrukkingen zijn voor stijl en oeuvre behoeft uitgebreide validatie.

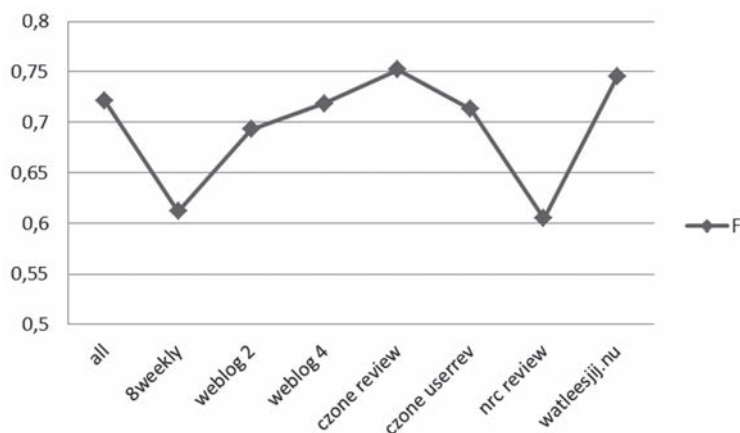
Grafiek 2 *Frequenties van positieve en negatieve emotiewoorden.*Grafiek 3 *Frequenties van verwijzingen naar ander werk en van stilistische terminologie.*

begin. Wat we hier zien is precies de bestaansreden van een corpus als dit – de ontwikkeling van betere tools vereist de beschikbaarheid van een grote collectie relevante teksten. Technieken voor *topic modelling*, zoals Latent Semantic Analysis (LSA), zouden hier ook een rol kunnen spelen (Louwerse en Van Peer 2007).

### *Bepalen van de strekking van een bespreking*

Iets anders wat we graag zouden willen is het automatisch bepalen van de strekking van een bespreking. Dat brengt ons op het terrein van *sentiment mining*. Ik heb het testcorpus geanalyseerd met behulp van het *sentiment mining tool* dat deel uit maakt van de *Pattern*-software, ontwikkeld door Tom Desmedt en Walter Daelemans (2012a). Pattern bepaalt de polariteit (de positieve of negatieve strekking) van een tekst voornamelijk op basis van de bijvoeglijke naamwoorden, en maakt daarvoor gebruik van een sentimentlexicon. Voor die teksten waarvoor de bespreker een aantal sterretjes aan de tekst heeft toegekend, kunnen we de geme-

Grafiek 4 F-score van door Pattern voorspelde positieve of negatieve rating.



ten polariteit vergelijken met het aantal sterren. Grafiek 4 geeft aan hoe goed de door Pattern voorspelde rating correspondeert met de werkelijk gegeven rating, voor alle sites en voor een aantal collecties afzonderlijk.<sup>27</sup> De score is niet slecht voor de betrekkelijk korte teksten, geschreven door ‘gebruikers’, zoals op watleesjij.nu, de Crimezone gebruikersbesprekingen en opvallend genoeg ook de ‘officiële’ besprekingen op de Crimezone-site. Bij langere en meer doorwrochte teksten, zoals die uit de NRC en van 8Weekly, is de score veel lager, waarschijnlijk vooral omdat er meer tekst aanwezig is die niet rechtstreeks gerelateerd is aan het eindoordeel over het boek. Aanpassingen in de algoritmes zouden deze scores zeker kunnen verbeteren (bijvoorbeeld door rekening te houden met de positie van woorden in de tekst, door rekening te houden met syntactische verbanden, door sarcasme te detecteren, etc.).

### Vooruitblik

Hoewel ze stuk voor stuk nadere uitwerking behoeven, kunnen de bovenstaande analyses nu al daadwerkelijk op het testcorpus worden uitgevoerd. Uitwerking zou moeten resulteren in hulpmiddelen (programmatuur) voor de analyse van online-boekdiscussie, die de gebruiker bij de bestudering van het corpus kan inzetten. Daarbij moet in het oog worden gehouden dat al deze analyses zijn gebaseerd op ‘niet-lezen’ (Kirschenbaum 2007), dus op het gebruik van statistische gegevens, en dat ze dus nooit de garantie geven dat individuele gevallen correct worden geïnterpreteerd.

De mogelijkheden van een corpus zoals hier beschreven reiken echter veel verder dan alleen de studie van online boekdiscussie alleen. Teksten over literatuur,

<sup>27</sup> Technisch: er wordt gekeken naar de groep besprekingen die een meer dan gemiddelde rating bevatten. Deze groep wordt vergeleken met de groep waarvan de door Pattern gemeten polariteit boven een bepaalde grenswaarde uitkomt. Van deze meting wordt de *recall* bepaald (welk deel van de groep met hoge rating is bevat in de groep met hoge polariteit) en de *precision* (welk deel van de groep met hoge polariteit is bevat in de groep met hoge rating). De kwaliteit van de voorspelling wordt gegeven door de F-score, het meetkundig gemiddelde van *recall* en *precision*.

vooral wanneer ze vergezeld gaan van een *rating*, kunnen ook geanalyseerd worden met het oog op de vraag wat mensen in literatuur waarderen. Zeker wanneer een analyse van de reacties op literatuur in verband kan worden gebracht met analyse van de literaire tekst zelf zou dat een aantal beperkingen van de sociologische blik op kwaliteitstoeschrijving aan literatuur kunnen opheffen. Dat is, vooralsnog, toekomstmuziek, maar wel een belangrijke reden om aandacht te besteden aan online boekdiscussie.

#### 4 Methodologische overwegingen

Terug naar de vraag wat deze vorm van onderzoek voor de discipline betekent. Een aantal zaken springt in het oog. Ten eerste, niet verrassend, de grote aantallen. Dankzij de computer wordt het, eindelijk, mogelijk om grotere hoeveelheden teksten in de beschouwing te betrekken, en dat is in elk geval voor een verschijnsel als online-boekdiscussie ook nodig. Er is vaak gezegd dat de geesteswetenschapper het unieke bestudeert, maar het unieke is onkenbaar. Kennis van teksten bestaat alleen in vergelijkende zin, bij de gratie van algemene patronen, op basis van overeenkomsten en verschillen, in verbanden waarvan je wilt kunnen vaststellen dat ze niet op toeval berusten, en kan alleen tot stand komen bij (eventueel impliciete) studie van grotere aantallen.

Ten tweede, de noodzaak voor het ontwikkelen van instrumenten. In het verleden kon het verzamelen van materiaal voor onderzoek, afhankelijk van het voorwerp van studie, een flinke klus zijn. Als materiaal digitaal beschikbaar is, zal die taak nu soms lichter worden, hoewel zeker niet altijd. Maar het conceptuele instrumentarium waarmee teksten vervolgens werden onderzocht kon in het verleden uit bestaande literatuur worden afgeleid en daarna toegepast. In de nieuwe situatie zal vaak gelden dat instrumenten moeten worden ontwikkeld, bijvoorbeeld om te meten wat een criticus belangrijk vindt. Er zijn technologieën die daarvoor kunnen worden ingezet, maar die technologieën zijn nog volop in ontwikkeling, ontwikkeld voor andere terreinen dan dat van de literatuurstudie, en vragen aanmerkelijke aanpassingen om bruikbaar te zijn voor letterkundige doeleinden.

Ten derde, en in samenhang daarmee, de werktuigen van de *digital humanities*-onderzoeker zijn nog bot. Ze kijken meestal naar woorden, vaak alleen de meest gebruikte woorden, en kijken niet naar hun volgorde. Als ze naar structuren kijken, kijken ze naar zinnen, niet naar het niveau van het betoog. En ze zijn altijd statistisch van aard: ze bepalen de betekenis van het geheel, min of meer als een optelsom van de delen – althans van die delen die zich laten tellen. De kans dat een computer de komende twintig jaar een reactie op een boek beter kan lezen dan u of ik dat kunnen is nihil.

Ten vierde, door de import in de literatuurstudie van methoden die elders zijn ontwikkeld, ontstaat een discontinuïteit in de traditie van het vak. We kunnen technieken die van elders komen, toepassen zonder ons noodzakelijkerwijs steeds af te vragen hoe ze zich verhouden tot wat onze voorgangers hebben gedacht en geschreven. Ik zie dat dat soms leidt tot verbazing bij onderzoekers met sterke wortels in de bestaande onderzoekstraditie. Met alle respect voor de traditie denk ik toch ook dat er, omdat er zo ongelofelijk veel is wat we nog niet weten, en nog



niet kunnen, alle reden is om ook onconventionele methoden een kans te geven. De literatuurwetenschap heeft de afgelopen eeuw vaker impulsen van buiten verwerkt, en is daar niet slechter van geworden.

Samenvattend: de *digital humanities* stellen ons in staat instrumenten te ontwikkelen waarmee we grote aantallen verschijnselen kunnen bestuderen en daarmee, op termijn, de traditionele voorkeur van de geesteswetenschappen voor het eenmalige en unieke kunnen verminderen. Tegelijkertijd moet worden erkend dat die instrumenten nog ernstige tekortkomingen hebben en dat het geenszins vast staat dat ze alles kunnen wat we ervan hopen. Daar staat tegenover dat de menselijke onderzoeker ook belangrijke beperkingen heeft: ons lezen gaat hopeloos traag en, veel belangrijker, we springen af op het eerste beetje betekenis dat we zien; we zijn makkelijk te overreden, en we overreden onszelf heel gemakkelijk, van het bestaan van twijfelachtige verbanden; en voor het waarnemen van tal van patronen zijn onze hersenen eenvoudig niet toegerust. Op het moment dat we machines leren om die patronen te herkennen, creëren we voor onszelf de mogelijkheid om meer materiaal en meer aspecten van dat materiaal te overzien.

Dat betekent niet noodzakelijk dat de computer ons definitieve antwoorden gaat geven op letterkundige vragen. Kirschenbaum laat zien hoe in het MONK-project, waar onderzoek werd gedaan naar het taalgebruik van Emily Dickinson, ‘while the machine learning algorithms are here iterating over several hundred texts, “reading” them at a distance, the end result is attention to individual words, the building blocks of language and poetry’. In dat proces is de machine niet een vervanging van, maar een aanvulling op het oordeelsvermogen en inzicht van de menselijke expert. Stephen Ramsay maakt in *Reading machines* (Ramsay 2011) hetzelfde punt: ‘algorithmic procedures can be made to conform to the methodological project of *inventio* without transforming the nature of computation or limiting the rhetorical range of critical inquiry’. Volgens Ramsay is dat het geval omdat elke kritische lezing van een tekst, niet alleen de computationele lezing, berust op een transformatie. Het toepassen van software voor die transformatie is de voortzetting van een bestaande praktijk. Het resultaat van die toepassing is een nieuwe tekst die nog steeds om interpretatie vraagt (Sculley & Pasanek 2008).

De toepassing van werktuigen is dus een uitbreiding van onze mogelijkheden, geen wondermiddel. Maar voor het werktuigen makende dier, dat de mens volgens Benjamin Franklin is, is de ontwikkeling van werktuigen om literatuur te analyseren en te begrijpen een logische stap. Als we die werktuigen vakkundig hanteren, zullen onze inzichten waardevoller zijn dan zonder die werktuigen.

## Bibliografie

- Absillis 2010 – K. Absillis, ‘Volksverheffing 2.0. Literatuurkritiek op het internet’. In: *Ons Erfdeel* 53 (2010) 2, p. 44-53.  
 Andringa 2006 – E. Andringa, ‘Penetrating the Dutch Polysystem. The Reception of Virginia Woolf, 1920-2000’. In: *Poetics Today* 27 (2006), p. 501-568.  
 Boot 2008 – P. Boot, ‘Te groot voor het blote oog. Over interactieve visualisatie in de studie van correspondenties’. In: *TNTL* 124 (2008) 3, p. 201-210.  
 Boot 2011 – P. Boot, ‘Towards a Genre Analysis of Online Book Discussion. Socializing, Participa-

- tion and Publication in the Dutch Booksphere'. In: *Selected Papers of Internet Research* IR 12.0 (2011).
- Boot 2012a – P. Boot, 'Online boekdiscussie: 8Weekly'. In: *textualscholarship.nl*. The Hague, 2012. <http://www.textualscholarship.nl/?p=11967>.
- Boot 2012b – P. Boot, 'Online boekdiscussie: watleesjij.nu'. In: *textualscholarship.nl*. The Hague, 2012. <http://www.textualscholarship.nl/?p=11522>
- Clement e.a. 2008 – T. Clement, S. Steger, J. Unsworth & K. Uszkalo, 'How Not to Read a Million Books'. In: *Seminar on the History of the Book*. New Brunswick, NJ: Rutgers University, 2008.
- Crane 2006 – G. Crane, 'What Do You Do with a Million Books?'. In: *D-Lib Magazine* 12 (2006) 3.
- Dera 2012 – J. Dera, 'Facebook als pr-strategie, debatplaats en expositieruimte. Over de rol van sociale media in het literaire veld'. In: *Ons Erfdeel* 2012, [http://www.onserfdeel.be/Jeroen\\_Dera\\_over\\_sociale\\_media\\_en\\_literatuur.pdf](http://www.onserfdeel.be/Jeroen_Dera_over_sociale_media_en_literatuur.pdf).
- Eder & Rybicki 2011 – M. Eder & J. Rybicki, 'Stylometry with R'. In: *Digital Humanities 2011: Conference Abstracts*. Stanford, CA: Stanford University, 2011.
- Joosten 2010 – J. H. T. Joosten, "'Toch is het niet altijd even hoogstaand en prijzenswaardig". Nederlandstalige literatuurkritiek op het internet'. In: *SDL* 52 (2010) 1, p. 85-108.
- Kirschenbaum 2007 – M. G. Kirschenbaum, 'The Remaking of Reading. Data Mining and the Digital Humanities'. In: *National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*. Baltimore, 2007.
- Koolen e.a. 2012 – M. Koolen, G. Kazai, J. Kamps, M. Preminger, A. Doucet & M. Landoni, 'Overview of the INEX 2012 social book search track'. In: *INEX 12 preproceedings*, <https://inex.mmci.uni-saarland.de/static/.../INEX2012-preproceedings.pdf>.
- Louwerse & Van Peer 2007 – M. Louwerse & W. Van Peer, 'Waar het over gaat in cijfers. LSA als kwantitatieve benadering in tekst- en literatuurwetenschap'. In: *TNTL* 123 (2007) 1, p. 22-36.
- McDonald 2007 – R. McDonald, *The Death of the Critic*. London, New York, 2007.
- Moretti 2005 – F. Moretti, *Graphs, Maps, Trees. Abstract Models for Literary History*. London [etc.], 2005.
- Moretti 2013a – F. Moretti, 'Conjectures on World Literature'. In: *Distant Reading*. London/New York, 2013, p. 43-62.
- Moretti 2013b – F. Moretti, 'Network Theory, Plot Analysis'. In: *Distant Reading*. London/New York, 2013, p. 211-240.
- Prescott 2012 – A. Prescott, 'Making the Digital Human. Anxieties, Possibilities, Challenges'. In: *Digital Humanities Summer School*. Oxford, 2012.
- Ramsay 2011 – S. Ramsay, *Reading Machines. Toward and Algorithmic Criticism, Topics in the digital humanities*. Urbana [etc.], 2011.
- Van Rees 1985 – K. Van Rees, 'Consensusvorming in de literatuurkritiek'. In: H. Verdaasdonk (ed.), *De regels van de smaak*. Amsterdam, 1985, p. 59-85.
- Sculley & Pasanek 2008 – D. Sculley & B. M. Pasanek, 'Meaning and Mining. The Impact of Implicit Assumptions in Data Mining for the Humanities'. In: *Lit Linguist Computing* 23 (2008) 4, p. 409-424.
- De Smedt & Daelemans 2012a – T. De Smedt & W. Daelemans, 'Pattern for Python'. In: *The Journal of Machine Learning Research* 13 (2012), p. 2031-2035.
- De Smedt & Daelemans 2012b – T. De Smedt & W. Daelemans, "'Vreselijk mooi!" (terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives'. Paper read at Proceedings of the 8th Language Resources and Evaluation Conference (LREC '12), 2012.
- Verboord 2011 – M. Verboord, 'Cultural Products go Online. Comparing the Internet and Print Media on Distributions of Gender, Genre and Commercial Success'. In: *Communications* 36 (2011) 4, p. 441-462.
- Verdaasdonk 2008a – H. Verdaasdonk, 'Analogieën als instrumenten om literaire fictie te classificeren'. In: J. Joosten, W. de Nooy en D. Verdaasdonk (ed.), *Snijvlakken van de literatuurwetenschap*. Nijmegen, 2008, p. 95-110.
- Verdaasdonk 2008b – H. Verdaasdonk, 'De kans om de P.C. Hooft-prijs te winnen'. In: J. Joosten, W. de Nooy & D. Verdaasdonk (ed.), *Snijvlakken van de literatuurwetenschap*, Nijmegen, 2008, p. 125-142.
- Verdaasdonk 2008c – H. Verdaasdonk, 'Hoe citaten het waardeoordeel van boekbesprekers weer geven'. In: J. Joosten, W. de Nooy & D. Verdaasdonk (ed.), *Snijvlakken van de literatuurwetenschap*. Nijmegen, 2008, p. 79-93.

Zijlstra e.a. 2004 – H. Zijlstra, T. van Meerveld, H. van Middendorp, J.W. Pennebaker & R. Geenen: 'De Nederlandse versie van de "Linguistic Inquiry and Word Count" (LIWC)'. In: *Gedrag & Gezondheid* 32 (2004), p. 271-281.

### Adres van de auteur

Huygens Instituut voor Nederlandse Geschiedenis KNAW  
Postbus 90754  
2509 LT Den Haag  
peter.boot@huygens.knaw.nl