

ELS STRONKS

De afstand tussen *close* en *distant*

Methoden en vraagstellingen in computationeel letterkundig onderzoek*

Abstract – In this article, the gap between *close* and *distant reading* methods and research design is discussed to establish if the two could profitably be combined. In the current situation, the use of *close* and *distant reading* techniques is not evenly distributed among literary scholars specializing in Dutch literature. A project like Nederlab aims to lower the threshold of technical expertise to overcome the reluctance to adopt *distant reading* techniques. What possibilities for *close reading* are opened up by digital technology?

De tegenstelling tussen *close* en *distant reading* die centraal staat in de oratie van Karina van Dalen-Oskam, roept associaties op van afstand en van verschil.¹ In deze bijdrage wil ik die associaties rationaliseren door ze uit te splitsen in twee categorieën. Enerzijds gaat het bij *close* en *distant reading* om verschillen in methoden, om twee verschillende manieren van lezen en analyseren die elkaar goed zouden kunnen aanvullen in plaats van uitsluiten. Anderzijds gaat het om een verschil in vraagstellingen. De introductie van *distant reading* is gepaard gegaan met de gedachte dat nieuwe vraagstellingen mogelijk, wenselijk of zelfs noodzakelijk zijn als de computer een deel van het analysewerk van de onderzoeker overneemt. Die gedachte vond weinig weerklank en schiep afstand tussen de traditionele vakbeoefening en *distant reading*. In recente toepassingen van *distant reading* lijken vooral traditionele vragen van een overtuigend antwoord te kunnen worden voorzien. Is dat de manier waarop de afstand tussen *close* en *distant* het beste te overbruggen is? Of zouden we toe kunnen werken naar een situatie waarin nieuwe digitale analysemethoden *close reading*-tradities op een andere manier aanvullen en verrijken?

1 De afstand in methoden

Als we onder *close reading* het minutieus lezen van teksten verstaan, kunnen we stellen dat deze methode onder letterkundigen al eeuwen beoefend wordt, vanuit wisselende inzichten. De humanisten deden aan *close reading* in de vroegmoderne periode met het idee dat zorgvuldig uitgevoerde filologische analyses hen dich-

* Met dank aan Peter Boot en de anonieme reviewers van een eerdere versie dit artikel.

¹ Zie voor de oratie http://oratiereeks.nl/upload/pdf/PDF-1290Weboratie_Van_Dalen-Oskam.pdf. In de oratie neemt Van Dalen-Oskam een aantal resultaten uit onderzoek naar stilistische aspecten van het werk van Betje Wolff en Aagje Deken die met *close reading* verkregen zijn als uitgangspunt, om te zien welke resultaten met *distant reading* in dat onderzoek geboekt kunnen worden.

terbij klassieke en bijbelse bronteksten zouden brengen. De oprichters van het tijdschrift *Merlyn* deden dat met het idee dat een analyse en interpretatie van de tekst alleen door lezing van die éne tekst tot stand kunnen komen. Het leggen van verbanden met auteursachtergronden en de context van die tekst (andere teksten, maatschappelijke omstandigheden) moest worden vermeden.² Anderen, zoals Edward Said, deden juist aan *close reading* om te laten zien dat die ene tekst ingebed was in discoursen in een samenleving. Analyseren en interpreteren heeft dan het doel verborgen maatschappelijke, ideologische agenda's op te sporen door de relatie tussen tekst en context aan te wijzen.³

Distant reading kunnen we pas sinds relatief kort, sinds teksten digitaal beschikbaar komen en we in grote corpora op zoek kunnen naar patronen en verbanden. Dat doen we door de computer woorden en woordcombinaties te laten indexeren, tellen en groeperen. *Distant reading* heeft iets van de truc met een potlood waarmee je over papier krast zodat voorheen onzichtbare zaken zichtbaar worden: we verwachten dat de computer met een andere leesmethode iets ongeziens aan ons zal vertonen. Het is veelzeggend dat een belangrijke verzamelplaats van digitale analysetools voor letterkundigen het volgende motto draagt: 'Voyant Tools: See Through Your Texts'.⁴

Die indruk wordt versterkt als je ziet hoe *distant reading* in de praktijk vaak werkt. Om grote verschuivingen in literaire stijlen op te sporen, lieten de Amerikaanse informatici Hughes, Foti, Krakauer en Rockmore een selectie uit het tekstcorpus van het Project Gutenberg door de computer analyseren op zogenaamde 'functiewoorden'. Dat zijn niet de woorden waar *close reading*-letterkundigen bij uitstek op letten (zelfstandige naamwoorden, werkwoorden) maar lidwoorden, tussenwerpsels en voegwoorden – 'content-free words', zeggen de onderzoekers zelfs, en die aanduiding alleen al zal bij sommigen een huivering oproepen. Maar aan het gebruik van die functiewoorden blijken verschuivingen in stijl af te meten.⁵ Conclusie van de onderzoeksgroep is dat na ca. 1900, literaire schrijvers elke 25 jaar van schrijfstijl veranderen. Voor die tijd zijn dergelijke stijlwisselingen niet aan de orde, en schrijven hele generaties op niet van elkaar te onderscheiden wijze: 'whereas authors of the 18th and 19th centuries continued to be influenced by previous centuries, authors of the late 20th century are strongly influenced by authors from their own decade' (Hughes e.a. 2012: 7685).

Op dit onderzoek kom ik nog terug als ik in het tweede deel van deze bijdrage op het verband tussen type vraagstellingen en *distant reading* inga. Voor nu constateer ik op basis van het onderzoek van Hughes en partners dat *distant reading* resultaten oplevert die met welke variant van de traditionele *close reading*-methode ook niet te verkrijgen zijn, omdat met het blote oog dergelijke hoeveelheden teksten niet te overzien zijn en ook niet gelezen kunnen worden zoals de computer ze leest.⁶ Dus is er reden beide analysemethoden te waarderen voor wat ze

2 Van Huizen 1962.

3 Karavanta & Morgan 2008: 9.

4 Zie <http://voyant-tools.org/>.

5 Een andere methode, gebaseerd op een 'classification algorithm' dat bekend staat als 'Nearest Shrunken Centroid' werd onlangs voor stilistische analyses van een corpus van 19^e-eeuwse romans beproefd door Jockers (Jockers 2013).

6 Om hun corpus te selecteren, namen de onderzoekers een set van 7.733 werken van auteurs van

waard zijn en ze samen te gebruiken om tot optimaal resultaat te komen. Dat betekent naar mijn idee dat we beide methoden niet moeten vermengen tot een ongedefinieerde mix van *close* en *distant reading*. We moeten de afstand tussen *distant* en *close* juist zo groot mogelijk maken: *distant reading*-technieken verder ontwikkelen en de technische problemen oplossen die er nu nog zijn, maar ook met nog meer overgave en precisie aan *close reading* doen om ook de vaardigheid daarvan verder te ontwikkelen, om nieuwe interpretaties van teksten te genereren en om de met *distant reading* verkregen resultaten goed te duiden.

Ik zie drie redenen waarom we beide methoden zouden moeten ontwikkelen en de tijd achter ons dienen te laten dat we de ene methode als een bedreiging van de ander zagen:

- 1 met *distant reading* kun je betekenissen, interpretatiemogelijkheden en patronen in teksten op het spoor komen die je met *close reading*-technieken niet vindt, en *vice versa*;
- 2 met *distant reading* kun je interpretaties controleren die met *close reading* verkregen zijn, en *vice versa*;
- 3 met *distant reading* kun je *close reading*-resultaten op representativiteit toetsen.

Veel overtuigende voorbeelden van opbrengsten van het type 1) zijn er nog niet, mogelijk omdat de afstand tussen *distant* en *close* vaak nog niet groot genoeg is.⁷ In het geval van stijlonderzoek (stylometrie) – waarvan het net genoemde onderzoek van de Amerikaanse informatici een voorbeeld is – is die afstand er al wel, omdat patronen in woorden worden geanalyseerd die letterkundigen normaliter negeren. Maar overtuigende toepassingen van *distant reading*-technieken in letterkundig onderzoek dat puur gericht is op het vergroten van inzicht in de betekenis van teksten, bleven tot op heden vrijwel uit omdat *distant reading* in dat onderzoek nog te weinig op de reken- en denkkracht van de computer berust en teveel op *close reading*-technieken.

Ik zal met een voorbeeldje uitleggen wat ik bedoel. Met Peter Boot ondernam ik eens een primitieve poging tot *distant reading* in een onderzoek naar de vermeende saaiheid van het werk van Jacob Cats. Naar alle waarschijnlijkheid was Cats de populairste, meest gelezen auteur van de Gouden Eeuw, maar als je aan *close reading* doet, is die aantrekkingskracht moeilijk te begrijpen.⁸ Wat dan opvalt is vooral de monotonie van zijn toon en zijn didactische aanpak. Hoe kon hij daarmee zoveel, zowel heel jonge als heel oude lezers bij de les houden? We keken met name naar Cats' *Sinne- en minnebeelden*. Die bundel werd in 1618 voor het eerst uitgegeven, in drie losse delen met liefdesemblemen voor jongeren, volwassenen en ouderen. Alleen het deel voor de jongeren verkocht heel goed, en dat noopte Cats om bij de herdruk van de bundel in 1627 alle delen samen te voegen, en in één

wie werk gedigitaliseerd is in het kader van het Project Gutenberg. Ze kozen werken geschreven na 1550, van auteurs die tenminste vijf werken in het Engels geschreven hadden (onduidelijk is of de onderzoekers ook vertalingen in het Engels meenamen in die set) en van wie geboorte- en sterfdatum bekend zijn. Zie: Hughes e.a. 2012: 7682.

⁷ Een overtuigend voorbeeld van dit type onderzoek is te vinden in Pennebaker & Stirman 2001. De auteurs onderzochten het woordgebruik van suïcidale dichters, en vonden dat dat significant anders was het woordgebruik van niet-suïcidale dichters.

⁸ Zie ook Kloek 1998.

band te koop aan te bieden. De samengestelde bundel verkocht daarna zo goed dat we mogen aannemen dat toch ook volwassenen en ouderen er wel iets van hun gading in vonden.⁹ We keken naar differentiatie in retorische technieken als aanspreekvormen (van ‘jij moet’ tot ‘zullen we’, en van ‘denk je ook niet dat’ tot ‘wij zullen’), de positie van Cats ten opzichte van de aangesprokene (van gelijke tot autoriteit), naar de opbouw van argumentatie, en de aard ervan (beloning of straf), en codeerden vormen die we tegenkwamen. Daarna lieten we de computer tellen. We kwamen erachter dat Cats in zijn retorische strategieën varieert al naar gelang de leeftijd van zijn doelgroep, en betoogden dat in die variatie misschien wel de verklaring voor zijn verkoopsucces schuilt. Tegenover de jeugd en ouderen is hij het meest persoonlijk, optredend in respectievelijk de ik- en wij-vorm. Het lyrisch ‘ik’ krijgt in de toepassingen voor de jongeren de ruimte om verzuchtingen te uiten, terwijl de ouderen door het gebruik van de wij-vorm bij de boodschap worden betrokken. De volwassenen laat Cats ruimte door ze onpersoonlijker te benaderen. In het merendeel van de gevallen dienen zij zelf te bepalen of ze zich aangesproken voelen door ze te lezen. Aansporen doet Cats de volwassenen door ze af te schrikken met negatieve beelden die zij op zichzelf dienen te betrekken. De bejaarden kunnen zich veelal spiegelen aan positieve voorbeelden. Hun wordt ook relatief vaak een beloning in het vooruitzicht gesteld als stimulans, terwijl de volwassenen de angstaanjagende consequenties van slecht gedrag worden voorgehouden.¹⁰

We vonden weliswaar iets dat we eerder niet zagen, maar de methode waarmee we dat deden was primitief omdat die was gebaseerd op *close reading*: we codeerden de verschillende retorische strategieën zelf, en al het interpretatiewerk kwam nog steeds van ons. We lieten de computer alleen tellen. Dat is het soort *distant reading* waar we niet heel veel wijzer van worden. We schreven het artikel over Cats in 2003, en het probleem waar we toen tegenaan liepen werd al veel eerder gesignaleerd en ter discussie gesteld,¹¹ maar een oplossing lijkt er inmiddels nog niet te zijn. Hoogstens zou je kunnen zeggen dat het idee dat bepaalde aspecten van teksten te tellen zijn, het besef met zich mee heeft gebracht dat resultaten van *close reading* controleerbaar kunnen worden gemaakt. Hoe er precies gecodeerd/geïnterpreteerd is, kan door geïnteresseerden in veel van dergelijk onderzoek tot op het niveau van de analyse van een enkel woord worden bekeken.¹²

Resultaten van het type 2), *distant reading* uitgevoerd om eerdere resultaten of vermoedens die uit *close reading* naar voren zijn gekomen te toetsen, zijn al wel in grote hoeveelheden voorhanden. Nemen we bijvoorbeeld de datering van Homerus’ *Ilias*: door *close reading*-classici is gesuggereerd dat dat epos in de achtste eeuw voor Christus ontstaan is, en die suggestie is onlangs bevestigd in grootschalig linguïstisch onderzoek.¹³ Zo zijn ook cultuurhistorici in staat gebleken om de wisselwerking tussen culturele veranderingen en de opkomst van bepaalde literai-

9 Luijten [2003], ‘Wanneer later het geheel in elkaar wordt geschoven, volgen de tweënvijftig emblemen elkaar in een zich herhalende drieslag op’.

10 Boot & Stronks 2003.

11 Zie bijvoorbeeld Stubbs 1995.

12 Bij het papieren artikel publiceerden we een webpagina waarop de coderingen te bekijken zijn, zie: <http://emblems.let.uu.nl/catsretorica/html/c1627searchaspect.html>

13 Altschuler e.a. 2013, gebaseerd op ‘evolutionary-linguistic phylogenetic statistical methods to differences in Homeric, Modern Greek and ancient Hittite vocabulary items’.

re genres te onderzoeken middels grootschalige semantische analyses. Onderzoekers constateerden bijvoorbeeld verbanden tussen de groeiende aandacht voor sociale interactie tussen hoofdpersonen in de opkomende 19^e-eeuws Engelse roman en de sociale veranderingen die zich op dat moment in de Engelse maatschappij voltrokken: de romans leverden een spiegel waarin lezers op dergelijke processen konden reflecteren, om vervolgens in het dagelijks leven aan vergelijkbare processen deel te nemen.¹⁴ Dichter bij huis leverde Mike Kestemont in zijn promotie-onderzoek overtuigende antwoorden op (deels reeds lang bestaande) vragen rond het auteurschap van een aantal Middelnederlandse teksten.¹⁵

Op dit terrein worden ingewikkelde linguïstische, semantische en semiotische analyses gemaakt die kennis van analysemethoden als *text mining* vereisen, maar zelfs met een veel simpelere techniek – het tellen van woordfrequenties – zijn al interessante resultaten te boeken. Een voorbeeldje van onderzoek dat ik kortgeleden uitprobeerde, in het kader van

het NWO-project Dutch Songs On Line: in dat project leveren we *full tekst*-versies van Nederlandstalige liederen, onder andere van de circa 40.000 liederen die er tussen 1500 en 1800 in de Nederlanden verschenen. In dat enorme corpus van 40.000 liederen valt het grote aantal liederen voor jongeren op (ik definieer dat even losjes als ‘jongeren in de huwbare leeftijd’). Wat stond er zoal in die liederen? Bijvoorbeeld een in onze ogen niet ongewone, mogelijk ironisch bedoelde conversatie tussen een moeder en dochter over te frivole kleding die aandacht van jongens dient te trekken. De moeder adviseert die kleding niet aan te doen, de dochter eindigt met de oproep aan andere meisjes dat juist wél te doen:

‘Een t’Samen-spraeck, tusschen Moeder en Dochter’
Stemme: Ick bender getrouwt.

Och Dochter hoe hangt’er u Borsjes soo bloot?
Men sietse tot aen de Tepel,
Schaemt ghy u niet, Dat men het siet,
De klokke hangt aen de klepel, la, la,
De klokke hangt aen de klepel.

Och moeder het is de nieuwe tijdt,
Ghy hebt het mijn noyt verboden,
Ick soeck’er een man, ick kander niet an,
Het isser de nieuwe mode, la, la,
Het isser de nieuwe mode.

[...]

Oorlof ghy Dochters in ’t gemeen
Wilt ghy de Vryers behagen,
Doet hier als ick, u lustigh op-schick,
Sy sullen u gunst toe-dragen, la, la,
Sy sullen u gunst toe-dragen.¹⁶

14 Zie Heuser & Le-Khac 2012.

15 Zie voor een bespreking en samenvatting van die resultaten de recensie van Kestemonts proefschrift van Joris van Zundert, <http://www.textualscholarship.nl/?p=12134>.

16 *Uytertse Hylickmaeckers* [1677]: 10-12.

Een lied als deze samenspraak tussen moeder en dochter kun je interpreteren als een middel om jongerenidentiteiten publiekelijk vorm te geven en ter discussie te stellen. Ik verwijs voor een uitgebreide uitleg van die gedachtegang naar een artikel dat ik onlangs schreef over Nederlandstalige studentenliederen. Het onderzoek waar dat artikel op gebaseerd is, liet zien dat in de vroege achttiende eeuw studenten zich al als soldaten gaan profileren (mogelijk om af te komen van hun imago als drinkebroers en gokkers), terwijl ze pas veel later in de maatschappelijke werkelijkheid ook echt soldaat worden en door de overheid in die rol worden geplaatst.¹⁷ Voor nu is van belang dat het bestaan van al die liedboekjes erop lijkt te wijzen dat jongeren die liederen lazen en zongen, aan de hand van die liederen bepaalden hoe ze zich konden gedragen. Tegelijkertijd hadden die liederen ook de functie de maatschappij een bepaald beeld van die jongeren te presenteren. Ze lijken het familiale, private leven te laten zien, maar hadden de functie de jongvolwassene – wiens identiteit op dat moment in de vroegmoderne maatschappij nog nauwelijks zichtbaar was – een publiek aanzien te geven.¹⁸ Men was zich van dergelijke processen van beeld- en imagovorming bewust, zo kunnen we aflezen uit een embleem uit Roemer Visschers *Sinne-poppen* uit 1614, getiteld ‘Ick gheeft haer weder’. Op de afbeelding zien we een jonge vrouw in een spiegel kijken. De begeleidende prozatekst legt uit:

DEN wereltlijcken handel [=gedrag] ende wandel is den aert van de Spieghel niet onghelijck: want in den Spieghel is niet dan een schijn, het dingh dat ghy daer in sien wilt, moet ghy selfs voor u brenghen: wat personagie ghy in de Wereldt spelen wilt, moet ghy in u selfs vormen.¹⁹

De lezer wordt middels zo’n embleem opgeroepen actief over zijn/haar rol en ‘personagie’ (‘een rol in een toneelstuk’, geeft het *WNT* hier als eerste betekenis) in de wereld na te denken. Het gedrag dat iemand tentoonspreidt, wordt vergeleken met een spiegelbeeld. Een rol bestaat pas als je die actief vormgeeft, zoals ook het spiegelbeeld pas verschijnt als iemand voor de spiegel gaat staan.

De hypothese van mijn onlangs gestarte onderzoek is dat in dat grote aanbod aan liederen de ons zo bekende identiteit van de rebellerende puber en jongvolwassene ontstond. Die identiteit kwam juist daar tot stand, omdat in die liederen zowel jongeren als volwassenen (ouders, kerkelijk leiders) zich in het discours mengden. Als we het hierboven geciteerde lied ironisch lezen, is aan te nemen dat het door een volwassene geschreven is die jongerengedrag ter discussie wil stellen. Dergelijke interactie tussen volwassenen en jongeren vinden we in die liedcultuur voortdurend.²⁰ Andere historische bronnen stonden ons tot op heden niet toe te zien hoe dat proces verliep, maar aan de liederen kunnen we aflezen hoe jongeren en volwassenen bouwden aan een identiteit die ergens in de periode 1500-1800 een maatschappelijke en sociale realiteit werd.

Mijn vermoeden, op basis van *close reading* verkregen, was dat het begin van de specifieke jongerenliedcultuur zich heel precies laat bepalen. In het *Antwerps liedboek* dat in 1544 verscheen, staan maar heel weinig liederen voor jongeren, en in

17 Stronks 2012.

18 Eisenbichler 2002.

19 Geciteerd via Adams 2009: 43.

20 Zie ook Veldhorst 2009 en Roberts 2012.

het eerste liedboek dat wel speciaal voor jongeren werd gedrukt, het *Aemstelredams Amoureux Liedboek* uit 1589, heel veel. Ik deed een simpele test naar woordfrequentie om te zien welke woorden in beide bundels het meest voorkomen. Het resultaat van die telling geef ik hier weer in een *word cloud* die gegenereerd is door de eerder genoemde Voyant-tool:



Onmiddellijk is te zien dat de woorden ‘ik’ en ‘mij’ in de bundel uit 1589 veel vaker voorkomen, en dat lijkt me een aanwijzing dat de liederen uit dat jaar specifiek de persoon/de identiteit van de (jonge) zanger adresseren. Die identiteit kregen jongeren voorgespiegeld in de fictieve wereld van het lied.

Wil dit resultaat echt overtuigen, dan zouden alle bundels uit de periode 1500-1800 aan een soortgelijke test onderworpen moeten worden, met aanvullende vragen: vallen bepaalde woorden in bepaalde bundels steeds vaker, en wat is de directe context van die woorden? Antwoord op die aanvullende vragen kunnen we niet door het tellen van woordfrequentie maar wel middels *text mining* verkrijgen: daarmee kunnen we niet alleen zoeken naar het voorkomen van bepaalde woorden, maar ook naar patronen en modellen, dat wil zeggen naar bijvoorbeeld verbanden tussen bepaalde woorden of zinnen, of complexe concepten die uit meer dan één woord zijn samengesteld. Hoe zag een typisch jongerenlied er in 1580 uit, en in 1677? Zijn er verschillen te constateren, en waar kunnen die toe herleid worden? Zolang spellingsvarianten en niet goed gedigitaliseerde teksten het indexereren (op basis van spellingsnormalisatie) van woorden belemmeren, kan dat *text mining* in historische corpora nog niet. Maar dat zijn bezwaren van technische aard, die verdere verkenning van het toetsen van *close reading*-resultaten middels *distant reading*-technieken niet in de weg moeten staan.²¹

Over nu naar resultaten van het type 3). De introductie van de methode van *distant reading* zal *close readers* mogelijk niet alleen nieuwe interpretaties en een toets van bestaande interpretaties opleveren, maar ook antwoorden op vragen over de representativiteit van vondsten. Letterkundigen die aan *close reading* doen, krijgen vaak – misschien niet zozeer van disciplinegenoten, maar wel van cultuurhistorici – de vraag: ‘wie zegt me dat jullie niet een paar pakkende voorbeelden van verschijnselen in teksten hebben gekozen, hoe vaak komen die voorbeelden eigenlijk voor?’ Als we er nu eens vanuit gaan dat we met *close reading*-technieken vaak speuren naar afwijkingen in patronen met het doel ideologische

21 Met dank aan José de Kruif, die voor mij de eerste experimenten met het corpus deed waaruit bleek welke hindernissen nog te nemen zijn.

discoursen in tekstcultuur te signaleren (herkenbaar aan ongemakkelijkheden, onregelmatigheden, frictie) dan kan *distant reading* ons helpen het gangbare patroon op te speuren. Met die methode kunnen we laten zien hoe vast een patroon was, hoe vaak afwijkingen voorkwamen, en onder welke condities. Mogelijk is het vooral overtuigend dat we kunnen laten zien hoe zeldzaam, en hoe afwijkend bepaalde teksten of elementen in teksten zijn. Het zou bijvoorbeeld bewijs kunnen leveren voor het idee dat bepaalde teksten hun makers toestaan te denken en dromen over zaken die er in andere delen van de tekstuele cultuur nog niet zijn.

Hier weer even een verwijzing naar dat jongerenliedonderzoek om deze toegevoegde waarde van het combineren van *close* en *distant reading* concreet te maken: kun je nu daadwerkelijk zien dat in die liedteksten andere tekstuele patronen gevonden worden dan in andere teksten uit die tijd, en is daarmee te onderbouwen dat zich in dat deel van de tekstuele cultuur iets anders afspeelde dan elders? Je zou grootschalig op zoek willen gaan naar verschillen in taalgebruik tussen liederen voor niet-jongeren en jongeren (ontstond er een eigen jargon?), maar ook naar tekstenmerken die in identiteitsvorming een rol spelen (aanspreekvormen, gebruik meervoud/enkelvoud, gebruik van bepaalde modi van werkwoorden die van de zangers bepaalde acties vroegen, relaties met anderen etc. etc.). Ook teksten in de context (kranten, pamfletten) zou je mee willen nemen, bijvoorbeeld om uit te zoeken of het opkomen van bepaalde thematiek te traceren is, of om te zien of bepaalde thematiek eerder in de liedcultuur dan elders opkwam.

2 De afstand in vraagstellingen

Ik heb het hier over een situatie die er nog niet echt is: als ik het eerlijk moet zeggen, zijn de letterkundigen nog meer van de *close* dan *distant reading*. Het is veelzeggend dat veel resultaten met *distant reading* door informatici geboekt worden. Kijk maar naar het genoemde onderzoek naar patronen in de wisseling van schrijfstijlen, naar Google *n-gram viewer*²² en naar onderzoek dat onder de noemer *culturomics* aan het ontstaan is.

Dat niet-letterkundigen de meeste initiatieven nemen, heeft een aantal gevolgen. Eén daarvan is dat de uitkomsten van onderzoek met *distant reading* gebrekkelijk geduid worden. De groep van Hughes deed onderzoek naar veranderingen in schrijfstijl, maar vond, zonder zich dat bewust te zijn, veel meer dan dat. Het onderzoek leverde ook bewijs voor de aanname dat in de vroegmoderne tijd het principe van de *imitatio* leidend was: dat principe verklaart waarom in de achttiende eeuw stijlwisselingen niet aan de orde waren. Men schreef met het doel eerdere voorbeelden te vervolmaken, en hield zich dus aan de schrijfstijl die al voorhanden was. Het is interessant te zien dat dit principe blijkens het onderzoek van de groep Hughes ook in de negentiende eeuw nog allesbepalend was. Het kan aan de streekproef liggen, maar feit is dat hierin nog niets te zien is van de opkomst van het principe van de originaliteit voor 1900. En ook dat is een uitkomst waar de onderzoekers geen betekenis aan geven, mogelijk omdat ze geen letterkundige naar hun resultaten hebben laten kijken.

²² Zie <http://books.google.com/ngrams>.

Gevolg van de dominantie van niet-letterkundigen in dit *distant reading*-onderzoek is ook dat we als letterkundigen niet meedenken over het type vraagstellingen dat met *distant reading* onderzocht wordt. Die vraagstellingen zijn momenteel voor het merendeel stilistisch dan wel cultuurhistorisch dan wel linguïstisch van aard. Nu valt met name dat stilistisch onderzoek voor sommige delen van de Nederlandse letterkunde gelijk met de *core business* van het vak – althans, volgens degenen die onlangs stelden dat voor de Middelnederlandse letterkunde het dateren van teksten sinds de negentiende eeuw ‘een van de belangrijkste uitdagingen’ (Kestemont & Caers 2011: 1) is. Maar voor veel andere perioden geldt dat niet: negentiende-eeuwse vragen naar stijl en datering zijn in vroegmodern en modern letterkundig onderzoek vervangen door vragen naar de esthetische en sociale functies van teksten. De ontwikkeling van een digitaal platform als Annotated Books Online, waarop vroegmoderne geannoteerde boeken onderzocht kunnen worden die ons interesseren omdat ze de receptie van teksten documenteren, is van die ontwikkeling een uitvloeisel.²³ Maar die ontwikkeling is in hedendaags *distant reading*-onderzoek nog niet of nauwelijks verdisconteerd.

We zullen actiever eigen vragen moeten inbrengen willen we daar verandering in brengen, en dat kan sinds kort door de inrichting van Nederlab, een recent gestart NWO-project dat tot doel heeft alle gedigitaliseerde Nederlandstalige teksten van ca. 800 tot heden op een platform samen te brengen, om die teksten met binnen Nederlab ontwikkelde, gebruiksvriendelijke tekstanalysesoftware te onderzoeken. Nederlab biedt hiermee een laboratorium voor letterkundig onderzoek – met name voor de historische letterkunde, moet ik erbij zeggen, omdat moderne teksten door *copyright issues* problematisch zijn – dat door de individuele onderzoeker niet zelf ingericht hoeft te worden.²⁴ Een nieuwe en veelbelovende oplossing voor het overbruggen van de afstand tussen letterkundigen en de computationeel onderzoek lijkt zich hiermee aangediend te hebben.

Bibliografie

- Adams 2009 – A.J. Adams, *Public Faces and Private Identities in Seventeenth-Century Holland: Portraiture and the Production of Community*. Cambridge: Cambridge University Press, 2009.
- Altschuler e.a. 2013 – E.L. Altschuler, A.S. Calude, A. Meade en M. Pagel, ‘Linguistic Evidence Supports Date for Homeric Epics’. In: *BioEssays* 2013, doi: 10.1002/bies.201200165.
- Anoniem z.j. – Anoniem, *Uytertse Hylickmaeckers*. Z.p. z.j. [Amsterdam: Cloppenburg, na 1677].
- Boot & Stronks 2003 – P. Boot & E. Stronks, ‘Ingrediënten van een succesformule. Digitaal onderzoek naar Cats’ *Sinne- en minnebeelden*’. In: *Nederlandse Letterkunde* 8 (2003) 1, p. 24–40.
- Eisenbichler 2002 – K. Eisenbichler, *The Premodern Teenager. Youth in Society 1150-1650*. Toronto: Toronto Centre for Reformation and Renaissance Studies, 2002.
- Grootes 1987 – E.K. Grootes, ‘Het jeugdig publiek van de “nieuwe liedboeken” in het eerste kwart van de zeventiende eeuw’. In: W. van den Berg & J. Stouten (red.), *Het woord aan de lezer. Zeven literatuurhistorische verkenningen*. Groningen: Wolters-Noordhoff, 1987, p. 72–88.
- Heuser & Le-Khac 2012 – R. Heuser & L. Le-Khac, ‘A Quantitative Literary History of 2,958 Nineteenth-Century British Novels. The Semantic Cohort Method’. In: *Literary Lab* 4 (2012),

²³ Zie www.annotatedbooksonline.com, een onderzoeksplatform van onderzoekers van universiteiten in Utrecht, Amsterdam, Princeton, Londen en Gent.

²⁴ Zie www.nederlab.nl, op dit moment is de eerste functionaliteit van de tools al aanwezig op de site.

- <http://litlab.stanford.edu/LiteraryLabPamphlet4.pdf>.
- Van Huizen 1962 – P.H. van Huizen, ‘De “simple didact” van het literaire specialisme? Een inleiding tot de kritiek van F.R. Leavis’. In: *Merlyn* 1 (1962-1963), p. 49-59.
- Houtsma 2009 – J. Houtsma, ‘Oude en nieuwe liedjes. Over de overlevering van populaire liederen in de zestiende, zeventiende en achttiende eeuw’. In: *Neerlandistiek.nl* (2009), <http://www.neerlandistiek.nl/publish/articles/000175/article.pdf>.
- Hughes e.a. 2012 – J.M. Hughes, N.J. Foti, D. Krakauer & D.N. Rockmore, ‘Quantitative patterns of stylistic influence in the evolution of literature’. In: *Proceedings of the National Academy of Sciences of the United States* 109 (2012) 20, p. 7682-7686.
- Jockers 2013 – M. Jockers, *Macroanalysis. Digital Methods and Literary Analysis*. Urbana/Chicago/Springfield: University of Illinois Press, 2013.
- Karavanta & Morgan 2008 – M. Karavanta & N. Morgan, *Edward Said and Jacques Derrida. Reconstellating Humanism and the Global Hybrid*. Newcastle: Cambridge Scholars Publishing, 2008.
- Kestemont & Caers 2011 – M. Kestemont & B. Caers, ‘Over de datering van de Middelnederlandse ridderepiek’. In: *Verslagen & Mededelingen van de KANTL* 121 (2011) 1, p. 1-59.
- Kloek 1998 – J. Kloek, ‘Burgerdeugd of burgermansdeugd? Het beeld van Jacob Cats als nationaal zedenmeester’. In: R. Aerts en H. te Velde (red.), *De stijl van de burger. Over Nederlandse burgerlijke cultuur vanaf de middeleeuwen*. Kampen: Kok Agora, 1998, p. 100-122.
- Luijten [2003] – J. Cats, *Sinne- en minnebeelden*. Ed. H. Luijten, Den Haag: SDU, 1996. Digitale versie 2003, zie <http://emblems.let.uu.nl/c1627.html>.
- Pennebaker & Stirman 2001 – J. Pennebaker & S.W. Stirman, ‘Word Use in the Poetry of Suicidal and Nonsuicidal Poets’. In: *Psychosomatic Medicine* 63 (2001) 4, p. 517-522.
- Roberts 2012 – B. Roberts, *Drugs before the Rock 'n Roll. Youth Culture and Masculinity in Holland's Golden Age*. Amsterdam: AUP, 2012.
- Stronks 2012 – E. Stronks, ‘Identiteiten van adolescenten in de vroegmoderne liedcultuur. Het studentenlied als casus’. In: *Nederlandse Letterkunde* 17 (2012) 3, p. 225-248.
- Stubbs 1995 – M. Stubbs, ‘Collocations and Semantic Profiles. On the Cause of the Trouble with Quantitative Studies’. In: *Functions of Language* 2 (1995) 1, p. 1-30, <http://www.lecschool.com.br/v1/biblioteca/ATUCorpora.pdf>.
- Veldhorst 2009 – N. Veldhorst, *Zingend door het leven. Het Nederlandse liedboek in de Gouden Eeuw*. Amsterdam: AUP, 2009.

Adres van de auteur

Universiteit Utrecht
 Trans 10
 3512 JK Utrecht
 e.stronks@uu.nl