

PETER BOOT

Te groot voor het blote oog

Over interactieve visualisatie in de studie van correspondenties

Abstract – The article argues for the importance of techniques for the visualisation and contextualisation of quantitative phenomena in collections of letters. Visualisation techniques support the ‘distant reading’ that Moretti has argued for in *Graphs, Maps and Trees*, which becomes increasingly important in an age of massive collections of digitised material. It is argued that good visualisations should be interactive, should provide access to the data they visualise, and should be addressable. Examples based on the correspondence of Hugo Grotius show how diagrams, maps and timelines can provide helpful context for studying the letters.

Zoals verschijnselen te klein kunnen zijn voor het blote oog – dan gebruiken we een microscoop om ze toch te kunnen bestuderen – of te ver – dan gebruiken we een telescoop – zo kunnen ze ook te groot zijn voor het blote oog, of te dichtbij. Dergelijke verschijnselen worden alleen maar zichtbaar door afstand te nemen, en dit artikel suggereert een aantal hulpmiddelen die daarbij kunnen helpen: wat we in die omstandigheden nodig hebben is een *megascoop*, een instrument om die grote verschijnselen zichtbaar te maken voor het blote oog en zo de studie ervan te vergemakkelijken.

Mijn uitgangspunt in dit artikel is de digitale beschikbaarheid van de bestudeerde teksten, in ons geval een correspondentie, maar dit artikel spreekt niet over de voordelen van doorzoekbaarheid, of wereldwijde beschikbaarheid, of de potentiële flexibiliteit van een digitale editie. Waar het nu om gaat is overzicht, en ik zal argumenteren dat overzicht tot stand komt door middel van contextualisatie en visualisatie.

Ik begin met een voorbeeld. De voorbeelden in dit artikel zijn ontleend aan de correspondentie van Hugo de Groot. De correspondentie, ongeveer 7500 brieven, is uitgegeven in zeventien dikke delen (Grotius 1928-2001). Stelt u zich voor dat u wilt weten hoe de correspondentie verliep tussen Grotius en Vondel, of Vossius. Als u gebruik maakt van de papieren editie, raadpleegt u daarvoor eerst de overkoepelende index in deel 17, ook als pdf beschikbaar, om te vinden in welke delen de betreffende correspondent voorkomt, vervolgens raadpleegt u de indices op de genoemde delen, en dan kunt u naar de betreffende brieven bladeren. Dat is een moeizaam proces, maar die constatering is niet bedoeld om de prestatie van de bezorgers van de brieven te kleineren, die met de mogelijkheden van het boek titanarbeid hebben verricht.

In een digitale editie kunnen we dat moeizame proces aanmerkelijk vereenvoudigen, en eenvoudigweg een correspondent uit een lijstje selecteren, zoals in figuur 1. De figuur is ontleend aan experimenten met een digitale versie van de correspondentie van De Groot, waar het Huygens Instituut op het ogenblik aan werkt. De figuren in dit artikel geven niet de uiteindelijke editie weer, het zijn voorbeelden die ons helpen na te denken over wat de mogelijkheden van de digitale editie moeten zijn. Ze zijn echter wel ‘echt’: het zijn geen schetsen van hoe een systeem

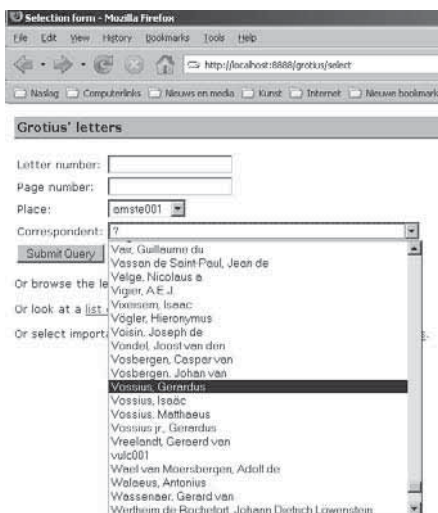


Fig. 1 (boven) Lijst van correspondenten van De Groot.

Fig. 2 (rechts) Overzicht van de correspondentie De Groot-Vossius.

Number	From	To	Date
295	Groot, Hugo de	Vossius, Gerardus	16131027
298	Groot, Hugo de	Vossius, Gerardus	16131101
299	Vossius, Gerardus	Groot, Hugo de	16131102
300	Groot, Hugo de	Vossius, Gerardus	16131102
323	Groot, Hugo de	Vossius, Gerardus	16140410
328	Vossius, Gerardus	Groot, Hugo de	16140414
332	Groot, Hugo de	Vossius, Gerardus	16140420
349	Groot, Hugo de	Vossius, Gerardus	16140612
354	Groot, Hugo de	Vossius, Gerardus	16140721
356	Groot, Hugo de	Vossius, Gerardus	16140801
359	Groot, Hugo de	Vossius, Gerardus	16140810
360	Vossius, Gerardus	Groot, Hugo de	16140811
362	Groot, Hugo de	Vossius, Gerardus	16140817
364	Groot, Hugo de	Vossius, Gerardus	16140821
366	Groot, Hugo de	Vossius, Gerardus	16140800
367	Groot, Hugo de	Vossius, Gerardus	16140800
368	Groot, Hugo de	Vossius, Gerardus	16140908
370	Groot, Hugo de	Vossius, Gerardus	16140914
372	Vossius, Gerardus	Groot, Hugo de	16140916
375	Groot, Hugo de	Vossius, Gerardus	16141011
376	Groot, Hugo de	Vossius, Gerardus	16141013
379	Groot, Hugo de	Vossius, Gerardus	16141020
381	Vossius, Gerardus	Groot, Hugo de	16141023
382	Groot, Hugo de	Vossius, Gerardus	16141023
383	Vossius, Gerardus	Groot, Hugo de	16141027
387	Groot, Hugo de	Vossius, Gerardus	16141221
390	Vossius, Gerardus	Groot, Hugo de	16150305
391	Groot, Hugo de	Vossius, Gerardus	16150311
392	Groot, Hugo de	Vossius, Gerardus	16150314

zou moeten werken, maar geven echte uitvoer van een systeem in ontwikkeling.

We kunnen dus een correspondent selecteren, en een overzicht krijgen van de brieven aan en van die correspondent (figuur 2). In dat overzicht zien we, zonder al te veel moeite, in welke jaren zich die correspondentie voornamelijk afspeelde. Voor een digitale editie zou dit voor zich moeten spreken. We kunnen echter nog een stapje verder gaan en een overzicht genereren als in figuur 3: een staafdiagram van de betreffende correspondentie, per jaar. Dit is een voorbeeld van wat ik een ‘megascoop’ noemde: een visuele weergave van een belangrijk gegeven over de correspondentie die in één oogopslag kan worden overzien.

Deze weergave leidt vanzelf tot een volgende vraag: wat is de positie van deze individuele correspondentie binnen het geheel van de correspondentie van De Groot? Om te beginnen zouden we het overzicht kunnen plaatsen in de periode van De Groots overgeleverde correspondentie (figuur 4). We krijgen dan een indruk in welke periode van zijn leven het contact met Vossius voor De Groot van belang was. Een betere indruk daarvan krijgen we misschien nog als we het relatieve volume van de correspondentie uitzetten (figuur 5): we zien hier voor elk jaar, in zwart, het percentage van de correspondentie van De Groot uit dat jaar, als deel van de totale correspondentie, en in grijs het percentage van de volledige correspondentie van De Groot met Vossius uit dat jaar. Wat we dus zien is een relatieve aanduiding van het belang van de correspondentie met Vossius in een bepaalde periode. We hebben nu een aspect van een correspondentie gevisualiseerd, en vervolgens hebben we het voorzien van een context in de vorm gegevens over de bredere correspondentie.

In plaats van te vragen naar een deelcorrespondentie en die te voorzien van context uit de bredere correspondentie, zouden we natuurlijk ook aan de andere kant kunnen beginnen: we kunnen beginnen te vragen naar de meest frequente corres-

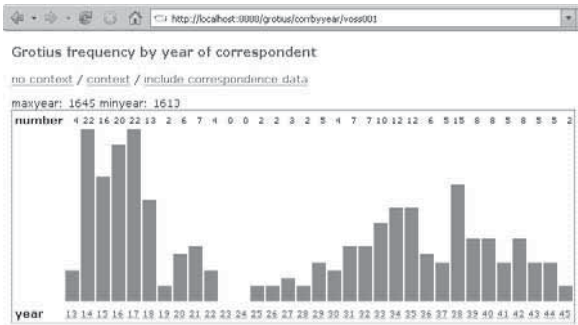
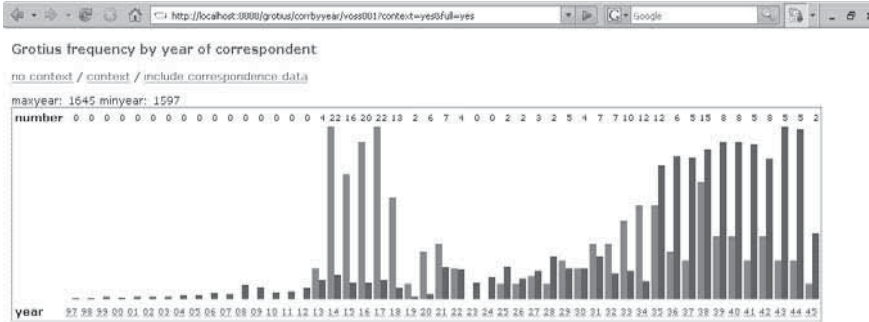
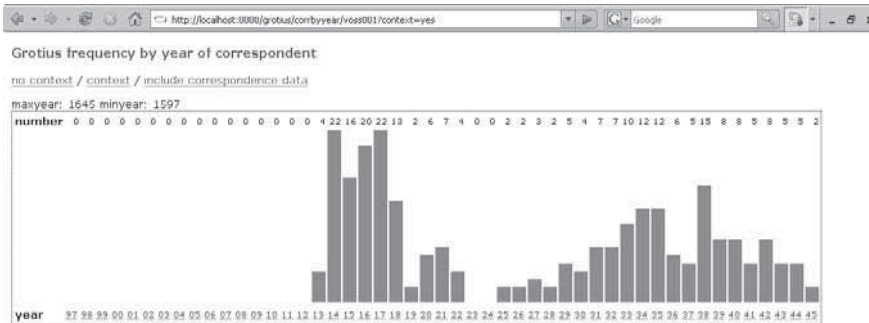


Fig. 3 (boven) De correspondentie De Groot-Vossius per jaar.

Fig. 4 (midden) De correspondentie De Groot-Vossius per jaar, geplaatst in de periode van De Groots overleverde correspondentie.

Fig. 5 (onder) De correspondentie De Groot-Vossius per jaar (grijs) ten opzichte van de Groots volledige correspondentie (zwart).



pondenten (figuur 6): de correspondenten gesorteerd op aantal brieven. Of, een andere manier om te kijken naar de belangrijke correspondenten, met wie strekt de correspondentie zich over de meeste jaren uit (figuur 7)? We zien de correspondenten en de duur van hun correspondentie met De Groot. Het zal niet verbazen dat dat vooral de familieleden zijn. De mogelijkheden zijn eindeloos. We zouden het belang van een correspondentie kunnen uitdrukken, niet in aantal brieven, maar in aantal lettertekens, zodat de lengte van de brief meeweegt. We zouden de weergave van de aantallen brieven kunnen afzetten tegen de weergave van de brieflengte, zodat we een indruk krijgen van aan wie een paar lange brieven zijn gericht, en aan wie talloze kattebelletjes. We kunnen de weergave beperken tot de brieven van De Groot zelf, in plaats van die van De Groot en correspondent. We kunnen de verschillende overzichten produceren per decennium of per



Fig. 6 De Groots frequente correspondenten (in aantallen brieven).



Fig. 7 De Groots langdurigste correspondenten (in aantal jaren).

jaar. The sky is the limit, of misschien geeft de diepte van de beurs de limiet aan, of misschien wel datgene wat nog overzichtelijk kan worden gepresenteerd – want dat is natuurlijk een volgende uitdaging.

Mijn stelling is dat grafieken zoals degene die we hier hebben gezien niet alleen maar aardig of handig zijn, maar een daadwerkelijk ander zicht op de historische werkelijkheid geven en ons dingen leren die in een andere presentatie van de gegevens onzichtbaar blijven.

Deze benadering is onder andere geïnspireerd door Franco Moretti's *Graphs, Maps, Trees* (Moretti 2005).¹ In het boek gebruikt Moretti grafieken, kaarten en bomen in de studie van de verschijnselen waar wij nu naar kijken, verschijnselen die te groot zijn voor het blote oog. Zijn keuze van technieken wordt mede gemotiveerd door de wens een literatuurgeschiedenis te schrijven die verder gaat dan onze beperkte canon: een literatuurgeschiedenis die rekening houdt met de literaturen van de hele wereld, en met de 99.5% van de werken die de canon niet halen. Het is duidelijk dat een dergelijke studie zich niet kan beperken tot technieken van 'close reading'. In zijn opstel 'Conjectures on world literature' (Moretti 2000) bepleit Moretti een methode van 'distant reading': 'where distance is however not an obstacle, but a specific form of knowledge: fewer elements, hence a sharper sense of their overall interconnection. Shapes, relations, structures. Forms. Models'.

Moretti's veld van onderzoek is de roman, en hij past grafieken toe om de ontwikkeling van subgenres van de roman uiteen te zetten, kaarten om, onder andere, patronen van gebeurtenissen in romans zichtbaar te maken – en vervolgens te relateren aan maatschappelijke ontwikkelingen – en bomen, vergelijkbaar met evolutionaire stambomen, om de totstandkoming van (sub)genrekenmerken te beschrijven. Hij beschrijft zijn toepassing van kaarten als volgt:

[...] you reduce the text to a few elements, and abstract them from the narrative flow, and construct a new, artificial object like the maps that I have been discussing. And with a little luck, these maps will be more than the sum of their parts: they will possess 'emerging' qualities, which were not visible at the lower level' (Moretti 2005:53).

Die 'emergente' eigenschappen, de eigenschappen die op een hoger niveau te voorschijn komen of zichtbaar worden, zijn de eigenschappen die ik omschreef als 'te groot voor het blote oog'. Een ander voorbeeld in ons geval zou kunnen gaan over de taal die in de brieven wordt gebruikt. Figuur 8 toont een weergave daarvan aan de hand van het eerste deel van Grotius' brieven: per correspondent een vak, waarvan de grootte correspondeert met het aantal brieven, en de kleur de taal weergeeft (lichtgrijs is Latijn, donkergrijs Nederlands, middengrijs Frans). Zo'n weergave roept direct de vraag op: wie zijn dan die paar mensen aan wie de Groot in het Nederlands schrijft, en waarom?

Moretti publiceert zijn grafieken en kaarten in een boek.² De boeken zijn de neerslag van heel veel geduldig onderzoek, en ze presenteren dat onderzoek aan

¹ Het zal niet zo vaak voorkomen dat tussen de geschenboekjes en de impulsaankopen op de toonbank van de boekhandel een Engelstalig boek over literatuurgeschiedenis ligt, maar, niet ten onrechte, viel onlangs het boek van Moretti die eer te beurt. Aan ontleding zal de wereld niet ten onder gaan.

² In eerste instantie in Moretti 1998.

Ten tweede: de grafieken als ingang tot de gegevens. De ondertitel van Moretti's boek luidt: 'abstract models for literary history'. Met die abstractie valt het wel mee, omdat uiteindelijk, zoals Moretti ook zegt, 'their consequences are on the other hand extremely concrete: graphs, maps, and trees place the literary field literally in front of our eyes'. Maar om de visualisaties te integreren in de praktijk van het onderzoek is het wenselijk ze niet alleen te kunnen zien, maar ze ook te gebruiken. Elk jaartal in een grafiek moet aanklikbaar zijn en leiden tot een overzicht van de brieven in dat jaar, elke persoonsnaam moet gekoppeld zijn aan de brieven aan die persoon, elke taal-persooncombinatie uit de grafiek die ik u zojuist toonde aan de brieven aan die persoon in die taal. Alleen dan worden de plaatjes inzetbaar voor het onderzoek van de vragen die ze oproepen.

Ten derde en ten slotte: de grafieken moeten bewaard en geadresseerd kunnen worden. Dat een afbeelding moet kunnen worden opgeslagen voor gebruik in een artikel spreekt voor zich. Maar met de blik op de toekomst is het misschien nog belangrijker om een bepaalde grafiek ook rechtstreeks digitaal te kunnen adresseren. Het opgeslagen plaatje is immers niet langer interactief en biedt geen toegang meer tot de gegevens. Als we een artikel schrijven waarin we gebruik maken van een bepaalde visualisatie, willen we kunnen verwijzen naar die interactieve versie die de achterliggende gegevens daadwerkelijk ontsluit. Dat betekent dat elke grafiek geïdentificeerd moet worden door een eigen webadres. Als dat niet het geval is, moeten we verwijzen in de vorm van ellenlange omschrijvingen als 'ga naar het hoofdmenu, selecteer Vossius als correspondent, vraag in de getoonde lijst om een presentatie van de correspondentie per jaar' – op papier misschien nog net acceptabel, maar als straks ook wetenschappelijke artikelen digitaal gaan verschijnen een echte belemmering voor effectief verwijzen.

Het Huygens Instituut werkt aan een database met geleerdencorrespondentie waarvan Grotius' correspondentie hopelijk een belangrijk onderdeel wordt.⁴ Het contextualiseren en visualiseren van gekwantificeerde informatie is één van de faciliteiten die we daarin willen gaan aanbieden. Zaken die dan mogelijk worden zijn bijvoorbeeld het opvragen van een kaart van de plaatsen van verzending en ontvangst van een bepaald epistolair subcorpus. Figuur 9 geeft een voorbeeld van het eerste deel van de brieven van De Groot. Na klikken op een plaats worden de aantallen getoond (figuur 10), vervolgens kunnen de betrokken brieven worden opgevraagd. De visualisatie is interactief en gekoppeld aan de data die worden gevisualiseerd.

Een andere optie is het tonen van een deel van de correspondentie op een tijdbalk, waarbij als context bijvoorbeeld een aantal feiten uit het leven van De Groot wordt getoond (figuur 11). Interactie is mogelijk door het schuiven van de balk, en opnieuw kunnen vanuit de visualisatie de onderliggende brieven worden opgevraagd (figuur 12). Nog weer andere mogelijkheden zijn het toepassen van hulpmiddelen voor geheel of gedeeltelijk automatische tekstclassificatie (Zie bijvoorbeeld Louwerse en Van Peer 2007), en het vervolgens visualiseren van de resultaten daarvan.

Het laatste punt dat ik hier aan de orde wil stellen heeft zijdelings te maken met

4 In het door NWO ondersteunde project 'Circulation of Knowledge and Learned Practices in the 17th century Dutch Republic'. In het project wordt samengewerkt met het Descartes Centre (Universiteit Utrecht), de KB, de Virtual Knowledge Studio (KNAW) en DANS.



Figuur 9 Plaatsen van verzending en ontvangst uit het eerste deel van De Groot's correspondentie.



Fig. 10 Aantallen brieven per plaats uit de correspondentie.

dat 'gedeeltelijk automatisch'. Figuur 9 toonde een kaart die met een druk op de knop door het systeem kan worden aangemaakt, een kaart van plaats van verzending en ontvangst in Grotius' correspondentie, maar helemaal vanzelf gaat dat

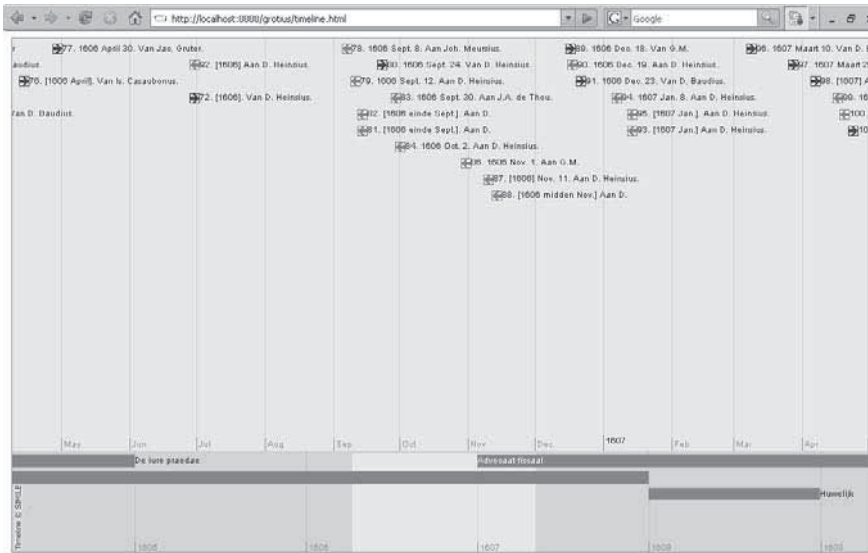


Fig. 11 De correspondentie van De Groot op een tijdbalk.

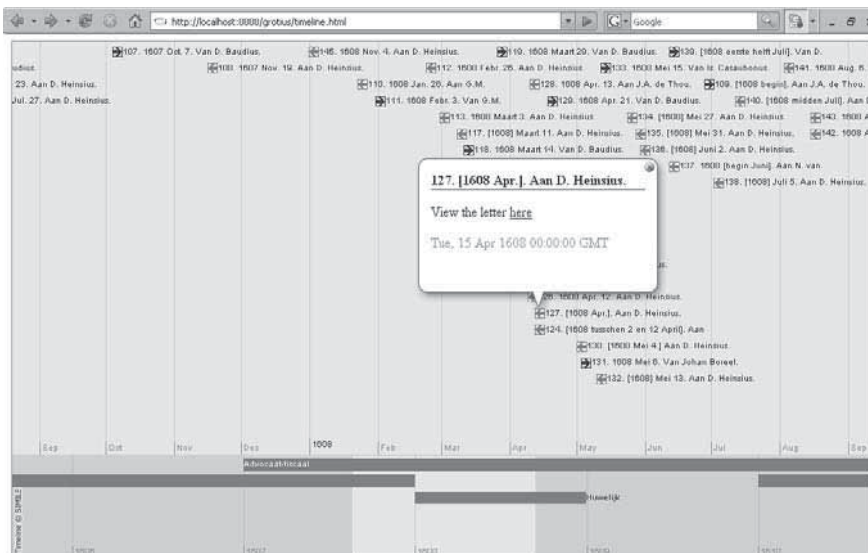


Fig. 12 Het opvragen van een brief uit de tijdbalk.

niet. Eerst moet iemand de plaatsen in de correspondentie hebben gecodeerd, en aan die plaatsen moeten geografische coördinaten zijn toegekend. Er komen in de correspondentie natuurlijk eindeloze hoeveelheden plaatsaanduidingen voor die nog niet expliciet zijn gecodeerd, en waarvan het toch interessant kan zijn ze op een kaart uit te zetten – denk aan woonadressen, plaatsen van herkomst, veldslagen, enzovoort. Behalve een component die visualisaties kan genereren aan de hand van reeds beschikbare gegevens, zullen onze edities het dus ook voor onder-

zoekers mogelijk moeten maken de betreffende gegevens in te voeren, zodat ze vervolgens de voor hun onderzoek relevante kaarten en grafieken kunnen tonen. Dat geldt niet alleen voor plaatsen: het zou mogelijk moeten worden om een categorisering naar onderwerp aan de brieven te hechten, of een groepering van de correspondenten in bijvoorbeeld familieleden, intellectuelen en diplomaten, of willekeurig welke andere onderscheiding die een onderzoeker relevant vindt. Dergelijke annotaties moeten net zo goed als basis voor visualisaties kunnen dienen als standaard aanwezige coderingen. Ook daarmee zijn we aan het experimenteren. Ik ga nog even terug naar het begin van mijn verhaal. Ik noemde de verschijnselen die te groot zijn voor het blote oog, en gebruikte die om de noodzaak van methoden voor visualisatie te beargumenteren. Zo lijkt het misschien alsof visualisatie-technieken alleen van belang zijn voor degenen die werken met zulke megacorpora als de brieven van De Groot. Ik denk dat dat een vergissing zou zijn. Elk corpus is een deelcorpus van een potentieel veel groter corpus van brieven uit het betreffende tijdvak, en de opmerkelijkheid van een verschijnsel kan alleen beoordeeld worden binnen dat grotere kader. Het zal om praktische redenen niet altijd mogelijk zijn, maar het is in principe wenselijk bij de studie van een kleinere correspondentie deze in te bedden in een grotere collectie van digitaal beschikbare en te analyseren brieven.

Wat ik hoop in deze bijdrage te hebben laten zien is dat interactieve visualisatie-technieken kwantitatieve verschijnselen in correspondenties niet alleen zichtbaar maar bijna tastbaar kunnen maken. Ze leveren een instrument voor de exploratie van de correspondentie dat in belangrijke mate kan bijdragen aan het overzicht over en inzicht in de correspondentie. Met de toenemende beschikbaarheid van grote digitale corpora wordt het experimenteren met en ontwikkelen van zulke interactieve visualisatietechnieken steeds belangrijker.

Bibliografie

- Grotius 1928-2001. – Hugo Grotius: *Briefwisseling van Hugo Grotius*. 17 delen. Den Haag (etc.), 1928-2001.
- Louwerse en Van Peer 2007. – Max Louwerse en Willie van Peer: 'Waar het over gaat in cijfers. LSA als kwantitatieve benadering in tekst- en literatuurwetenschap'. In: *TNTL* 123 (2007), p. 22-36.
- Moretti 1998. – Franco Moretti: *Atlas of the European Novel. 1800-1900*. London, 1998.
- Moretti 2000. – Franco Moretti: 'Conjectures on world literature'. In: *New Left Review* 1 (2000), p. 54-68.
- Moretti 2005. – Franco Moretti: *Graphs, maps, trees. Abstract models for literary history*. London (etc.), 2005.

Adres van de auteur

Huygens Instituut KNAW
 Postbus 90754
 NL-2509 LT Den Haag
 peter.boot@huygensinstituut.knaw.nl