

Karina van Dalen-Oskam

## Lezen van een afstand: de stijl van *Herfsttij der Middeleeuwen*

[karina.van.dalen@huygens.knaw.nl](mailto:karina.van.dalen@huygens.knaw.nl)

<https://www.huygens.knaw.nl/medewerkers/karina-van-dalen-oskam/>

Uitgeschreven versie van de lezing gehouden op het symposium

## Honderd jaar Herfsttij der Middeleeuwen

Een eerbetoon aan het bekendste werk van historicus Johan Huizinga

19 november 2019

### Inleiding

Mijn vakgebied is de computationele literatuurwetenschap. Een onderdeel daarvan is de stilometrie; daarin meten we teksten op allerlei manieren door. We tellen woorden en passen statistische methoden toe. Dat klinkt misschien heel saai, maar we kunnen er verrassende dingen mee doen. Zo konden we aantonen dat de onbekende debuterende romanauteur Robert Galbraith, van *The Cuckoo's Calling*, dezelfde persoon was als J.K. Rowling, auteur van de wereldberoemde Harry Potter-serie. En konden we twijfelaars aan Harper Lee als auteur van *To Kill a Mockingbird* geruststellen toen de oerversie van deze roman, *Go Set a Watchman*, verscheen – nu was eindelijk duidelijk dat zij wel degelijk zelf verantwoordelijk was voor de roman, die volgens kwade tongen alleen maar zo goed was geworden door de hulp van Truman Capote. En dan misschien nog het gevoeligste voorbeeld: Elena Ferrante. Met hulp van stilometrie is aangetoond dat de stijl van Ferrante's romans opvallend overeenkomt met die van de stijl van de romans van Domenico Starnone. Maar helemaal tevreden zijn we nog niet – er wordt aangenomen dat ook Anita Raja, vertaalster van onder meer Christa Wolf in het Italiaans, een grote invloed op de inhoud van de romans heeft gehad, maar dat is niet goed te meten omdat zij geen eigen romans heeft gepubliceerd.

Stilometrische auteursherkenning houdt een statistische analyse in van de hoogstfrequente woorden: auteurs hebben hun eigen 'ritme' in het gebruik van juist veelvoorkomende woorden. Dat zijn meestal functiewoorden als lidwoorden, persoonlijk voornaamwoorden, voorzetsels, enzovoorts. En daar zijn er genoeg van om statistisch overtuigende observaties te doen.

Maar er is meer mogelijk dan alleen auteursherkenning. Onderzoek naar schrijfstijl en vergelijking met grotere aantallen teksten en auteurs dan je met het blote oog en handmatig zou kunnen doen. Hoe verhouden genres zich ten opzichte van elkaar? Kunnen we verschillen zien tussen boeken die door lezers hoogliterair worden gevonden of minder literair? Hebben bestsellers een ander woordgebruik dan boeken die minder goed worden verkocht?

Dit soort onderzoek heet wel ‘distant reading’: lezen van een afstand. Onderzoekers lezen zoveel mogelijk van de teksten nog steeds zelf – geen zorgen! Maar we kunnen er ook nog heel veel andere boeken bij betrekken. En daar dan die titels in vinden die we absoluut ook nog moeten lezen voor ons onderzoek (en ons plezier).

Wat ik het mooie van deze methode vind – maar ik weet dat niet iedereen dat met me eens is – is dat de observaties in principe controleerbaar en herhaalbaar zijn. In principe, want dat kan alleen in die gevallen waarin de teksten aan anderen beschikbaar gemaakt mogen worden – en dat is lang niet altijd mogelijk. Andere onderzoekers kunnen in principe constateren of we correct hebben gemeten, en vervolgens kunnen we in discussie over de interpretatie. Met een veel concretere basis dan daarvoor.

Mij is gevraagd om voor deze gelegenheid *Herfsttij* van een afstand te lezen om een indruk te geven van de stijl van het werk. Onder stijl versta ik niet alleen wat traditioneel daaronder wordt verstaan, maar ook het gebruik van die hoogstfrequente woorden. Ook die hebben invloed op de stilistische uitstraling van een tekst.

Daarnaast neem ik ook afstand in de tijd. Ik vergelijk *Herfsttij* met tien hedendaagse non-fictie boeken waarin de Middeleeuwen of de Vroegmoderne tijd centraal staan en met tien hedendaagse literaire romans. Mijn doel is om een indruk te krijgen van de plaats die *Herfsttij* ten opzichte van die twintig boeken inneemt. Lijkt het woordgebruik meer op dat van de literaire romans of op dat van de non-fictie? Of zijn de Nederlandse taal en de conventies voor de genres waar we het over hebben zo veranderd dat een vergelijking niet zo zinnig is?

Dit is een klein experiment, dat alleen maar iets zegt over de in totaal 21 boeken die ik erin betrek. Maar toch geeft het een aardig beeld van hoe *Herfsttij* nu door lezers ervaren zou kunnen worden, en van de methoden die hedendaagse literatuurwetenschappers tot hun beschikking hebben.

## **De woordenschat van *Herfsttij***

Eerst een korte blik op de honderd meestfrequente woorden uit *Herfsttij*. Het meest voorkomende woord is het lidwoord *de*, dat 10.782 maal is aangetroffen. Op rang 10 staat *is*, een vorm van het koppelwerkwoord *zijn*, dat 1.522 keer is geteld. Dat is al een stuk minder vaak. Zo werkt dat voor alle teksten. En altijd weer verrassend om te zien: grofweg de helft van alle verschillende woorden in een tekst komt maar eenmaal voor.

Toch zijn er alleen al uit de honderd meestfrequente woorden aardige dingen af te leiden over *Herfsttij*. We zien dat er ook een aantal Franse functiewoorden in de top-100 staan (*et, la, le, que, ...*). Dat bevestigt het grote belang van het Frans in de teksten waaruit Huizinga citeerde. (Het kan trouwens zijn dat er onder de 10.782 voorkomens van *de* behalve het Nederlandse lidwoord *de* ook gevallen van het Franse voorzetsel *de* zitten en dat ook de vorm *des* uit beide talen komt.)

Verder is het interessant om vast te stellen welke zelfstandig naamwoorden in de top-100 in *Herfsttij* voorkomen: *leven, eeuw, tijd, geest, kunst, vorm, liefde, God...* Ook die geven al een inkijkje in de thema's van het werk.

De 100 meest frequente woorden uit <i>Herfsttij</i>					67	197		
1	10782	de	34	475	maar	68	195	que
2	5414	van	35	469	nog	69	190	kan
3	4966	het	36	420	zo	70	188	des
4	4442	en	37	417	ook	71	186	ons
5	3196	in	38	406	heeft	72	175	alles
6	3061	een	39	406	hem	73	170	zonder
7	1837	die	40	384	hun	74	168	l
8	1666	zijn	41	353	al	75	167	waren
9	1645	te	42	346	la	76	161	tegen
10	1522	is	43	336	had	77	160	kunst
11	1253	dat	44	330	naar	78	159	qui
12	1211	der	45	304	leven	79	159	werd
13	1102	met	46	303	wordt	80	159	à
14	1061	als	47	296	deze	81	157	s
15	1000	hij	48	283	ik	82	157	vorm
16	940	niet	49	279	geen	83	154	wij
17	871	aan	50	261	le	84	152	les
18	796	op	51	255	meer	85	152	liefde
19	754	voor	52	254	alle	86	152	tussen
20	681	om	53	253	zelf	87	150	want
21	665	was	54	248	hebben	88	147	zal
22	657	zij	55	246	wat	89	145	andere
23	643	men	56	238	worden	90	144	onder
24	599	et	57	233	over	91	142	nu
25	590	zich	58	230	d	92	142	wanneer
26	554	dan	59	222	eeuw	93	140	moet
27	551	of	60	220	grote	94	139	middeleeuwse
28	547	door	61	219	hoe	95	139	reeds
29	543	er	62	217	dit	96	137	veel
30	536	haar	63	215	zou	97	137	wel
31	530	bij	64	214	waar	98	136	ne
32	529	tot	65	207	tijd	99	136	weer
33	516	uit	66	198	geest	100	135	god

## Non-fictie

“Hoe individueel die stijlproeven ons wellicht voorkomen, de stijl van Huizinga was uit de aard der zaak tot zekere hoogte een historisch verschijnsel, toegankelijk voor een stijlhistorische analyse.”

F. Jansonius, *De stijl van Huizinga*

Voor mijn vergelijking van *Herfsttij* met hedendaagse non-fictie met een historisch of literatuurhistorisch onderwerp mocht ik gebruikmaken van de versie in hedendaagse spelling uit de jubileumeditie uitgegeven door Anton van der Lem. Ik benaderde zeven auteurs van

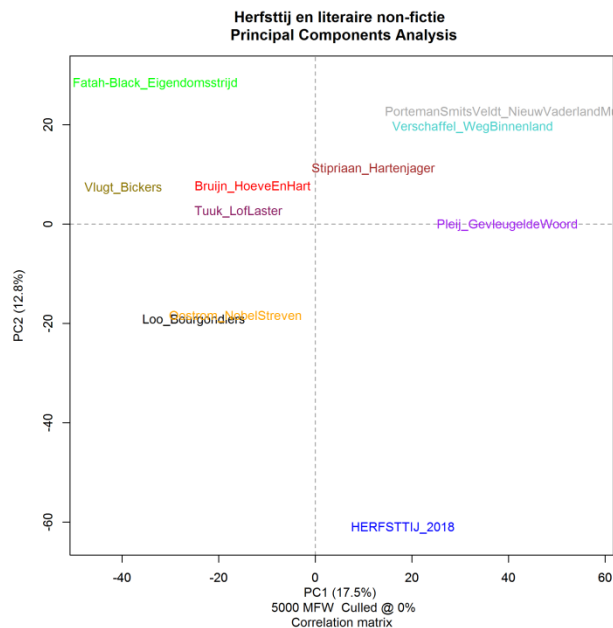
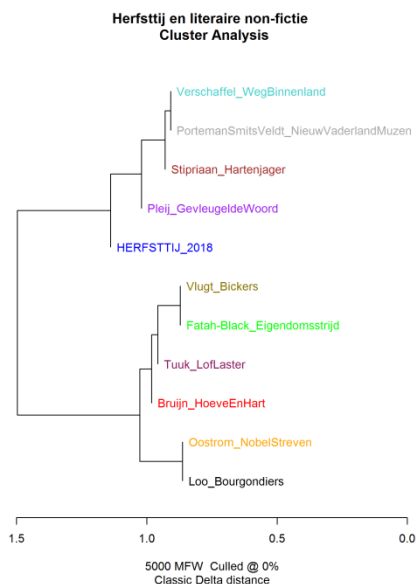
recente boeken over de Middeleeuwen en de Vroegmoderne tijd en al deze auteurs en hun uitgevers waren zo vriendelijk mij een digitale versie van het boek te bezorgen die ik kon gebruiken om de beoogde metingen te doen – dank hiervoor! Het gaat om de volgende zeven boeken:

Enny de Bruijn	<i>De hoeve en het hart</i>	(Prometheus)
Karwan Fatah-Black	<i>Eigendomsstrijd</i>	(Ambo   Anthos)
Bart Van Loo	<i>De Bourgondiërs</i>	(De Bezige Bij)
Frits van Oostrom	<i>Nobel Streven</i>	(Prometheus)
René van Stipriaan	<i>De hartenjager</i>	(Querido)
Luit van der Tuuk	<i>Lof en laster</i>	(Omniboek)
Simone van der Vlugt	<i>Wij zijn de Bickers!</i>	(Prometheus)

Tijdens de voorbereidingen voor de lezing werd de Digitale Bibliotheek voor de Nederlandse Letteren ([www.dbnl.org](http://www.dbnl.org)) uitgebreid met alle delen van de recentste literatuurgeschiedenis (gepubliceerd door uitgeverij Bert Bakker). Daaruit voegde ik nog drie titels toe aan mijn non-fictiecorpus:

Herman Pleij	<i>Het gevleugelde woord</i>
Karel Porteman en Mieke Smits-Veldt	<i>Een nieuw vaderland voor de muzen</i>
Tom Verschaffel	<i>De weg naar het binnenland</i>

Op deze tien teksten plus *Herfsttij* in de versie uit 2018 heb ik verschillende meetmethoden losgelaten. Hieronder laat ik de resultaten zien voor de 5.000 meestfrequente woorden, dus de eerste 5.000 woorden uit een lijst waarin alle voorkomende woorden in alle elf boeken samen staan gerangschikt met het meestfrequente woord op rang 1.

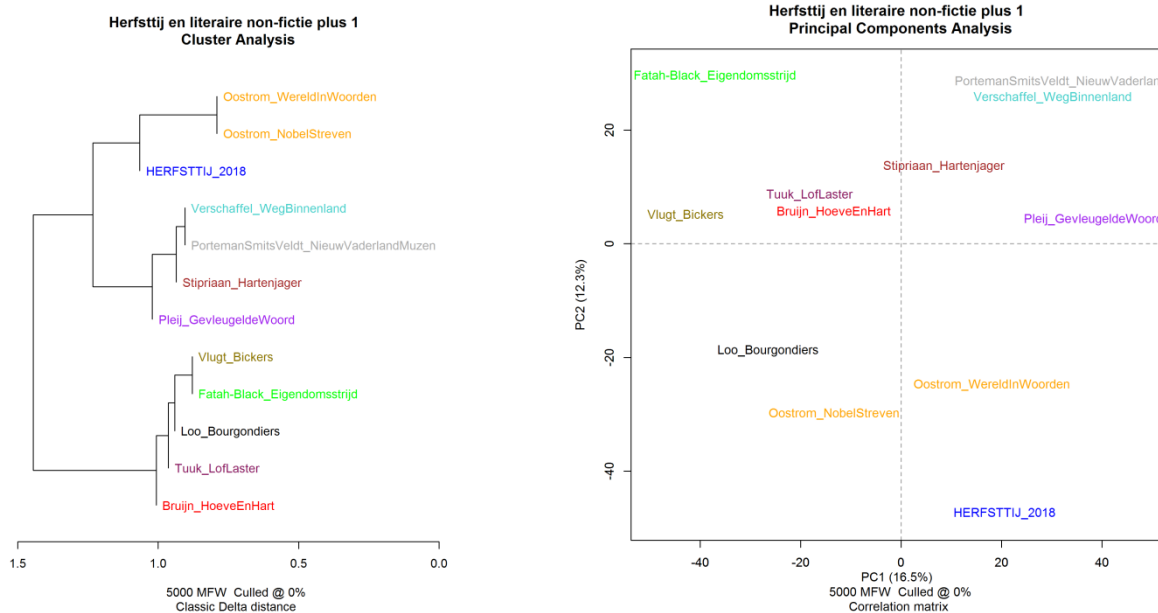


Links staat een visualisatie van een clusteranalyse. Hoe verder naar rechts twee titels door een verticale lijn met elkaar verbonden zijn, hoe meer ze wat het frequentiepatroon van die 5.000 meestfrequente woorden op elkaar lijken. Deze clustering leverde uiteindelijk twee hoofdclusters op, een van vijf en een van zes boeken. *Herfsttij* zien we in het cluster met de boeken van Pleij, Van Stipriaan, Porteman en Smits-Veldt, en Verschaffel. Het andere cluster bevat de boeken van Van Oostrom en Van Loo (die het meest op elkaar lijken in het woordgebruik dat is gemeten), Fatah-Black, Van der Vlugt, Van der Tuuk en De Bruijn.

Rechts staat een visualisatie van een andere meting, om na te gaan of die hetzelfde beeld oplevert – we vertrouwen nooit op maar één methode. De visualisatie is gebaseerd op een analyse op hoofdcomponenten. Wat deze analyse laat zien is dat *Herfsttij* in de eerste component (horizontaal afgebeeld) wel past binnen het non-fictiecorpus, maar dat het op de tweede component (afgebeeld op de verticale as) behoorlijk ver weg staat van alle andere titels. Dit kon de clusteranalyse niet goed duidelijk maken. Zo'n andere meting nuanceert de eerdere meting dus, en voorkomt dat we voorbarige conclusies trekken.

Metingen die ik beperkte tot de 1.000 meestfrequente woorden en metingen waarin ik alleen die woorden in de analyse betrok die in alle elf teksten voorkomen, leverden een vergelijkbaar beeld op. Dat wijst er op dat de inhoud en de onderwerpen van een boek niet de allergrootste invloed hebben op de resultaten en dat we dus wel degelijk een reflectie van de stijl van de betreffende auteur in de resultaten zien.

Zoals ik in de inleiding al zei, zeggen deze analyses alleen iets over de relatie van precies deze teksten ten opzichte van elkaar. Zodra er teksten worden weggehaald of toegevoegd, verandert het beeld. Dat kan ik illustreren door een tweede titel van een van de auteurs aan de vergelijking toe te voegen. Ik heb gekozen voor een van de twee delen uit de literatuurgeschiedenis die zijn geschreven door mijn promotor Frits van Oostrom.



De visualisaties van de clusteranalyse en van de analyse op hoofdcomponenten laten drie dingen zien. Ten eerste illustreren ze de kracht van stilometrie in het herkennen van auteurs, want de twee titels van Van Oostrom worden in de clusteranalyse gemarkeerd als het meest op elkaar lijkend. Ook in de meting op hoofdcomponenten staan ze bij elkaar in de buurt, maar iets minder dicht dan we wellicht hadden verwacht. Kern is dat er geen titels van anderen tussen de twee titels in staan.

Ten tweede zien we in de visualisatie van de clusteranalyse dat Van Oostrom nu verhuisd is naar het cluster met *Herfsttij*, en dat zijn twee boeken samen zelfs het meest op *Herfsttij* lijken en daarmee een subclustertje vormen.

Ten derde zien we in de visualisatie van de analyse op hoofdcomponenten dat op de tweede component (afgebeeld op de verticale as) het gat tussen de tien titels en *Herfsttij* nu kleiner geworden is en dat de boeken van Van Oostrom nu aanzienlijk dichterbij *Herfsttij* staan dan in de vorige meting, waarin het boek van Van Loo nog net ietsje dichterbij te vinden was.

Als ik een tweede boek van Van Loo zou toevoegen, kan dit beeld echter zo weer veranderen. Kortom: het is goed om nogmaals in gedachten te brengen dat we steeds alleen de gekozen teksten met elkaar vergelijken – en dat de relaties niet voor eeuwig en definitief vastliggen.

De volgende vraag is dan uiteraard: waarin verschillen deze boeken nu precies van elkaar? Ik doe daarvoor een iets andere meting. Ik vergelijk de woordenschat van *Herfsttij* met die van de tien non-fictieboeken en bereken welke woorden relatief gezien significant vaker voorkomen in *Herfsttij* en in de andere non-fictie. Ik som hieronder uit de top-100 alleen de zelfstandig naamwoorden op:

#### **Significant vaker in *Herfsttij* dan in de andere non-fictie**

*litteratuur, kunst, geest, gedachte, schoonheid, element, ridderideaal, renaissance, levensvorm, voorstelling, symbolisme, dingen, ideaal, Middeleeuwen, ...*

#### **Significant vaker in de andere non-fictie dan in *Herfsttij***

*teksten, literatuur, publiek, jaar, tekst, vrouwen, stad, auteur, bundel, rederijders, familie, manier, periode, toneel, republiek, jaren, auteurs, liederen, gedichten, succes, kamers, predikant, boeren, ...*

Ook met deze meting kunnen we verschillen tussen *Herfsttij* en de andere boeken zichtbaar en tastbaar maken. Het opvallendst is dat de spelling voor 'literatuur' systematisch verschilt: *Herfsttij* gebruikt *litteratuur* met twee t's, de andere auteurs houden het bij één t. *Herfsttij* heeft het verder significant vaker over bijvoorbeeld *kunst, schoonheid* en *ideaal*, terwijl de andere boeken significant vaker melding maken van *teksten, familie* en *succes*, om maar een paar voorbeelden te noemen. Zo'n benadering kan onderzoekers op het spoor zetten van aspecten die tijdens lezing met het blote oog nog niet waren opgevallen.

Laten we nu kijken naar de relatie tussen *Herfsttij* en hedendaagse literaire fictie.

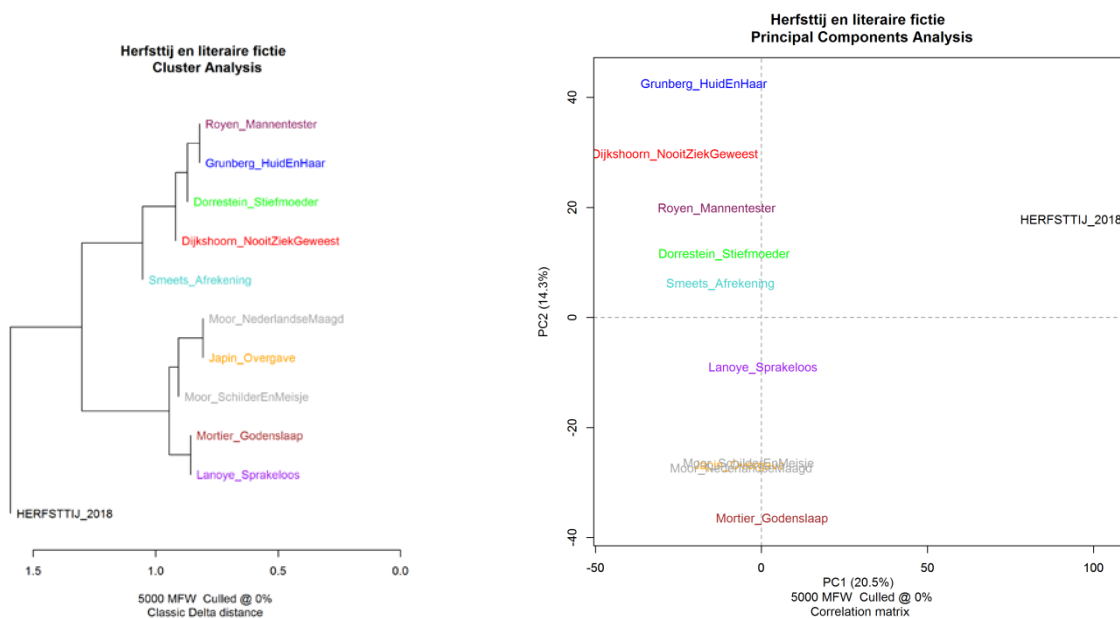
## Literaire fictie

“Johan Huizinga, onze beroemdste historicus, laat zich inderdaad veel beter vergelijken met schrijvers dan met andere historici.”  
Willem Otterspeer, *Orde en trouw: over Johan Huizinga*

Voor mijn vergelijking van *Herfsttij* met hedendaagse literaire fictie heb ik de volgende tien romans gebruikt.

Nico Dijkshoorn	<i>Nooit ziek geweest</i>
Renate Dorrestein	<i>De stiefmoeder</i>
Arnon Grunberg	<i>Huid en haar</i>
Arthur Japin	<i>De overgave</i>
Tom Lanoye	<i>Sprakeloos</i>
Marente de Moor	<i>De Nederlandse maagd</i>
Margriet de Moor	<i>De schilder en het meisje</i>
Erwin Mortier	<i>Godenslaap</i>
Heleen van Royen	<i>De mannentester</i>
Mart Smeets	<i>De afrekening</i>

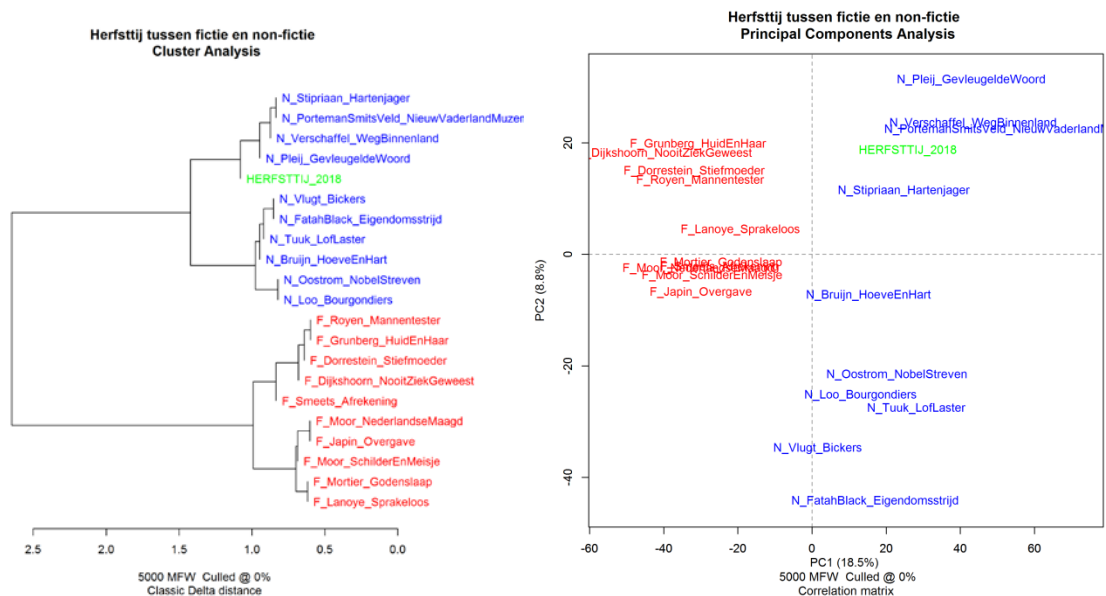
Een heel gevarieerde set van romans die door lezers heel verschillend worden gewaardeerd wat hun literaire waarde betreft (over deze verschillen in toegekende literaire waarde hoop ik volgend jaar meer te publiceren). Wat als we *Herfsttij* op dezelfde manier met deze romans vergelijken als hierboven met de non-fictie?



Het resultaat is duidelijk. In de visualisatie van de clusteranalyse (links) zien we dat *Herfsttij* een geheel eigen cluster inneemt, en dus aanzienlijk verschilt van de tien romans. De visualisatie van de analyse op hoofdcomponenten (rechts) geeft nog wat meer diepte aan de observatie: *Herfsttij* verschilt inderdaad het meest significant op de eerste component, afgebeeld op de horizontale as.

## Herfsttij tussen fictie en non-fictie

Hierboven heb ik uitgelegd dat het beeld bij elke weglating of toevoeging weer kan veranderen. De volgende stap is dan om *Herfsttij* met zowel de tien romans als met de non-fictie tegelijk te vergelijken. Waar vinden we Huizinga's beroemde werk dan terug?



Ook dit resultaat is glashelder. De twee tekstgenres zijn overtuigend van elkaar gescheiden in de visualisaties van de twee verschillende metingen, en *Herfsttij* blijft binnen de non-fictiegroep op dezelfde plaats staan die het eerder had toen ik het met alleen de non-fictietitels vergeleek. Kortom: het is zeer onwaarschijnlijk dat moderne lezers Johan Huizinga's *Herfsttij der Middeleeuwen* ervaren als een roman en niet als een non-fictieboek.

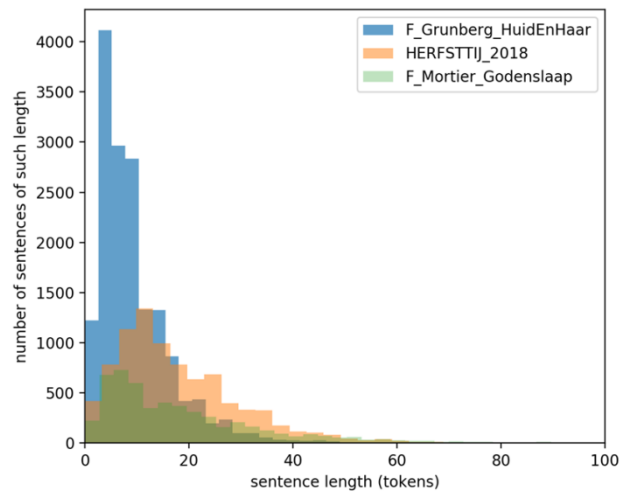
Ook op andere manieren sluit *Herfsttij* aan bij de hedendaagse non-fictieboeken die ik heb geselecteerd. Ik kon dankzij de hulp van mijn collega Joris van Zundert (Huygens ING) naar twee aspecten kijken: gemiddelde zinslengte en woordenrijkdom (omvang van de woordenschat, ook wel lexicale rijkdom genoemd). Ik ga hier alleen in op de zinslengte. Eerst de cijfers voor de gemiddelde zinslengte voor elke titel.



Zinslengte	Gemiddeld aantal woorden	Standaard-deviatie	Variatie
N Porteman & Smits-Veldt <i>Een nieuw vaderland</i>	20.0	11.0	120.5
N Fatah-Black <i>Eigendomsstrijd</i>	19.6	8.8	78.1
N Verschaffel <i>De weg naar het binnenland</i>	19.4	11.7	137.7
N Pleij <i>Het gevleugelde woord</i>	19.3	9.2	84.9
F Mortier <i>Godenslaap</i>	19.2	17.3	298.8
N van der Tuuk <i>Lof en laster</i>	18.9	8.3	69.3
N de Bruijn <i>De hoeve en het hart</i>	18.4	11.3	127.0
N van Oostrom <i>Nobel streven</i>	18.3	10.4	108.8
<b>HERFSTTIJ_2018</b>	<b>18.0</b>	<b>12.6</b>	<b>159.9</b>
N Van Loo <i>De Bourgondiërs</i>	17.4	10.1	101.5
N Stipriaan <i>De hartenjager</i>	16.1	10.4	107.9
N van der Vlugt <i>Wij zijn de Bickers!</i>	15.9	7.8	61.5
F de Moor <i>De schilder en het meisje</i>	15.4	11.9	141.6
F Japin <i>De overgave</i>	15.0	11.2	124.8
F Lanoye <i>Sprakeloos</i>	14.1	13.3	176.5
F Smeets <i>De afrekening</i>	13.9	9.7	94.9
F de Moor <i>De Nederlandse maagd</i>	12.1	8.4	69.7
F Dorrestein <i>De stiefmoeder</i>	11.1	8.2	67.0
F van Royen <i>De mannestester</i>	10.7	7.6	58.0
F Grunberg <i>Huid en haar</i>	9.6	7.5	55.5
F Dijkshoorn <i>Nooit ziek geweest</i>	8.3	5.7	32.6

De boeken zijn gerangschikt van hoogste gemiddelde zinslengte (aantal woorden per zin) naar laagste. Wat we hier zien is opnieuw een duidelijke scheiding tussen de twee tekstgenres. De non-fictie (in blauw, met een N voor de titel) groept samen aan de top – inclusief *Herfsttij* – en de fictie (in rood, met een F) aan de onderkant van de lijst. Met één uitzondering: de roman *Godenslaap* van Erwin Mortier heeft zich tussen de non-fictie genesteld. Dat is een interessante observatie, en kan inspireren tot een nadere analyse van de stijl van deze roman vanuit het perspectief van genreconventies voor fictie en non-fictie.

In de tabel hierboven is ook de standaarddeviatie en de variatie opgenomen. Deze twee kolommen laten zien dat alleen de gemiddelde zinslengte maar erg weinig zegt over een tekst als geheel. Wat ik daarom wilde weten was hoeveel zinnen van verschillende lengtes er in de 21 teksten uit mijn corpus voorkomen. Die vraag leidde tot de visualisatie op de volgende bladzijde. We zien hier het absolute aantal zinnen van een bepaalde lengte in *Herfsttij*, in de roman *Huid en haar* van Arnon Grunberg en in de roman *Godenslaap* van Erwin Mortier. Horizontaal het aantal woorden per zin, en verticaal het aantal zinnen van die lengte. Het boek van Grunberg stond bijna onderaan de ranglijst op gemiddelde zinslengte, met gemiddeld 9.6 woord per zin, en dat van Mortier stond heel hoog in de lijst met gemiddeld 19.2 woorden per zin. Huizinga's *Herfsttij* stond een paar plaatsen onder *Godenslaap* met gemiddeld 18 woorden per zin. Wat de visualisatie laat zien is dat *Huid en haar* relatief weinig variatie in zinslengte heeft: het blauw verspreid zich over een kleinere set aan zinslengten dan de andere twee boeken. Meer variatie in zinslengte laat *Herfsttij* (roze) zien, en *Godenslaap* (groen) wint het van deze drie in zinslengtevariatie.



## Conclusie

Het kleine experiment waarover ik heb gerapporteerd laat alleen observaties toe over de 21 boeken die in de analyse zijn meegenomen. Iets algemeen is er niet uit te concluderen. Wel heeft het een paar interessante observaties opgeleverd over genreverschillen tussen hedendaagse fictie en non-fictie en is het duidelijk dat *Herfsttij* in hedendaagse spelling zonder twijfel bij de non-fictie hoort.

Ik heb mijn experiment alleen op moderne teksten toegepast. Ik had heel graag nagegaan hoe de eerste editie van *Herfsttij* uit 1919 past in het toenmalige fictie- en non-fictielandschap. Kunnen we meten welke stilistische invloeden *Herfsttij* in 1919 vertoont? Om die vraag te beantwoorden zou ik het boek willen vergelijken met nu nog steeds bekende werken van historici en letterkundigen van toen, maar ook met titels die volledig uit zicht zijn verdwenen. Maar juist die laatste zijn nog niet in digitale vorm voorhanden. Ook is het een project op zich om een relevante set van boeken voor de vergelijking te kiezen.

Maar wie weet is er een onderzoeker die hierin een mooi onderwerp voor promotieonderzoek ziet, en kunnen we over een aantal jaren alsnog nagaan welke literaire en historische invloeden terug te vinden zijn in de stijl van Huizinga's *Herfsttij der Middeleeuwen*. De methoden zijn voorhanden.