

Hoe begrijpelijk is Van Maanen...?

Significant minder dan de Hoge Raad!

Willem van Boom¹

Deze bijdrage is opgenomen in: R. de Groot e.a. (red), *Kritiek op recht – Liber amicorum Gerrit van Maanen*, Deventer: Kluwer 2014, p. 43-60

(de originele paginanummering staat in **[vet]** weergegeven in de tekst)

1. Inleiding

[43] Met het emeritaat van Gerrit van Maanen – een van de meest sympathieke en tegelijk spraakmakende civilisten van de afgelopen jaren, wiens artikelen tamelijk vaak ingezonden reacties uitlokken en die van zijn hart geen moordkuil maakt – is het moment daar om terug te kijken op het werk dat Gerrit in zijn zeer vruchtbare academische carrière heeft geproduceerd. Natuurlijk kunnen we dan als jurist een inhoudelijke analyse van het werk van Gerrit maken, we kunnen de ontwikkeling in zijn werk van naïef Marxist tot landbezittende bourgeois in kaart brengen, we kunnen zelfs inhoudelijk het debat met hem aangaan. Deze bundel bevat ongetwijfeld meerdere van dergelijke bijdragen – het werk van Gerrit geeft meer dan genoeg aanleiding tot debat en tegenspraak. In deze bijdrage wil ik iets anders doen. Ik stel mij de vraag hoe begrijpelijk de teksten van Van Maanen eigenlijk zijn. Dat is een moeilijke vraag, want hoe kan men de begrijpelijkheid van tekst meten? Tegelijk is de vraag naar begrijpelijkheid in algemene zin wel uiterst relevant: academici die hun beroep **[44]** maken van het schrijven van tekst (zoals juristen), zullen willen weten of hun publiek de tekst begrijpt. Maar ook beleidsmakers zijn soms geïnteresseerd in begrijpelijkheid van geschreven tekst. De AFM wil bijvoorbeeld dat productfolders en algemene voorwaarden in de bank- en verzekeringsbranche ‘begrijpelijk’ zijn. Dat streven veronderstelt dat we weten hoe we begrijpelijkheid van tekst kunnen onderzoeken en ‘meten’. In deze bijdrage verken ik deze materie. Bovendien kan ik een belangrijke conclusie rapporteren: de Hoge Raad is 16-30% begrijpelijker dan Van Maanen.

¹ Deze bijdrage bouwt voort op mijn bijdrage ‘Begrijpelijke hypotheekvoorwaarden en consumentengedrag’ aan de bundel ‘Perspectieven voor vastgoedfinanciering’, Stichting Fundatie Bachiene 2014, en op M.R. van Dam, W.H. van Boom, M.L. Tuil, ‘Tekstbegrip en klantbelang bij financiële producten’, in: E.M. Dieben & F.M.A. ‘t Hart (red.), *Klantbelang Centraal* (Financieel Juridische Reeks 4), Amsterdam: NIBE-SVV 2012, p. 109 e.v. Ik ben veel dank verschuldigd aan het team student-assistenten dat de datavoorbereiding verzorgde. Bij wijze van ‘partes pro toto’ noem ik Sarah van Os en Jente Waal. Ik maakte dankbaar gebruik van nuttig commentaar op eerdere versies van Vincent Buskens, Pieter Desmet, Siewert Lindenbergh en Chris Reinders Folmer.

2. Meten is (iets meer) weten

2.1. Tekstniveau

Taalbeheersing zegt iets over personen, tekstniveau iets over geschreven tekst. Tekstniveau zou men kunnen onderscheiden in leesbaarheid en begrijpelijkheid.² De leesbaarheid wordt bijvoorbeeld beïnvloed door lettertype en –grootte, contrast, regelafstand enzovoorts. Begrijpelijkheid betreft vooral de moeilijkheidsgraad van een tekst, waarbij bijvoorbeeld lengte van zinnen, constructie en alledaagsheid van woorden een rol spelen. In teksten van Van Maanen valt bijvoorbeeld op het veelvuldig gebruik van liggende streepjes om nadere toelichting en uitweiding mogelijk te maken. Ik kan me voorstellen dat dergelijk gebruik invloed heeft op leesbaarheid én begrijpelijkheid. Begrijpelijkheid is een lastig te operationaliseren begrip, omdat de ervaring van moeilijkheid iets kan zeggen over het taalbeheersingsniveau van de lezer, maar ook omdat verschillende aspecten van een tekst in onderlinge samenhang van invloed zijn op de moeilijkheidsgraad.³ Zo brengen korte zinnen het tekstniveau omlaag. Korte zinnen met woorden die niemand kent, stuwen het [45] tekstniveau echter weer omhoog. Er zijn dus wel richtlijnen die iets over de invloed op het tekstniveau kunnen zeggen, maar hoe kun je die nu in onderling verband toepassen?

Daarover buigen taalwetenschappers zich. Er zijn in binnen- en vooral buitenland tal van manieren ontwikkeld om begrijpelijkheid van tekst te modelleren, dat wil zeggen in min of meer objectieve termen te vatten. Eenvoudig is dat allerm minst. Zo wordt ‘begrijpelijkheid’ vaak gezien als het resultaat van een interactie tussen lezer en tekst. Voor de schrijver van de tekst zelf is begrijpelijkheid iets anders dan voor de ongetrainde lezer, en voor mensen met een beperkte taligheid zal een tekst minder snel ‘begrijpelijk’ zijn dan voor mensen met een hoge mate van taligheid. Men zal dus een manier moeten ontwikkelen om vaststellingen over ‘begrijpelijkheid’ te objectiveren en zodoende herhaalbaar te maken. Bovendien is ‘begrijpelijkheid’ een *construct*, een niet direct waarneembaar fenomeen dat men door middel van een samenstel van wél waarneembare fenomenen moet proberen te modelleren.

Er bestaan verschillende modellen van begrijpelijkheid van tekst. Zo wordt voor onze taal wel gebruik gemaakt van het Europese Referentiekader (ERK)⁴ waarbij tekstniveaus zijn onderverdeeld in A1, A2, B1, B2, C1 en C2.⁵ De uitersten A1 en C2 duiden respectievelijk eenvoudige en zeer moeilijke tekstniveau's aan. Het nadeel van dit soort modellen voor het meten van tekstniveau is echter dat ze de schijn van alomvattendheid wekken, terwijl in werkelijkheid de achterliggende methodes niet per se transparant of gelijkloidend zijn. Metingen kunnen daardoor uiteenlopen en/of voor buitenstaanders ondoorgrondelijk lijken.⁶

² Overigens worden verschillende termen gebruikt om ‘tekstniveau’ aan te duiden, zoals begrijpelijkheid, moeilijkheidsgraad en dergelijke termen meer.

³ Vgl. over de verschillende manieren om tekstniveau te onderzoeken bijv. Van Dam et al. 2012.

⁴ Common European Framework of Reference for Languages: Learning, Teaching, Assessment.

⁵ Let wel: het ERK als zodanig is geen methode maar slechts een schaal. Nader L. Lentz, *Let op: Begrip verplicht! - Begrijpelijkheid als norm in de wet (oratie Utrecht)*, Utrecht: 2011, p. 1 e.v. Vgl. Van Dam et al 2012.

⁶ Vgl. C. Jansen, *Voorkomen is beter - Op weg naar effectievere teksten in de gezondheidscommunicatie (oratie Groningen)*, Groningen: 2012, p. 13 e.v.

[46] Een alternatief is het gebruik van meer objectieve en transparante ‘tel en turf’ modellen. Voor de Nederlandse taal is er bijvoorbeeld de *Flesch Douma Reading Ease* (FDRE) schaal. De FDRE is een schaal van 0 tot 120 die door een getalsmatige weging van gemiddelde zinslengte en het gemiddeld aantal lettergrepen per woord komt tot een cijferscore van tekstniveau.⁷ Hoe hoger de score, des te eenvoudiger het tekstniveau is. De gedachte hierachter is dus dat korte zinnen met woorden die uit weinig lettergrepen bestaan, eenvoudiger leesbaar zijn dan lange zinnen met woorden met veel lettergrepen. Een voorbeeld uit het oeuvre van Van Maanen kan dit verduidelijken:

Voorbeeldzin	FDRE score
<i>‘Soms komt op hem de verplichting te rusten om schadebeperkende maatregelen te nemen, bijvoorbeeld door een financiële bijdrage in de kosten van een zuiveringsinstallatie, zoals in het geval van de Kalimijnen.’ (G.E. van Maanen, NTBR 1994, p. 89)</i>	26,48
<i>‘Soms heeft hij de plicht om schade te beperken. Dat kan door financieel bij te dragen aan de kosten van het zuiveren van water. Dat gebeurde in de zaak van de Kalimijnen.’ (door mij herschreven versie)</i>	86,22

Slecht nieuws dus voor de lezer die het originele artikel van Van Maanen heeft gelezen, althans die betreffende zin. Een score van 26 op een schaal van 120 is belabberd. De door mij herschreven versie eronder is al een stuk eenvoudiger!

Uiteraard moet men zich niet blind staren op de FDRE score van een tekst. Het is een manier om tekstniveau eenvoudiger te maken, maar een hoge FDRE score is geen garantie voor begrijpelijkheid. Korte zinnen vol met korte woorden scoren namelijk prima op de F(D)RE schaal, maar als ze vol staan met niet alledaagse technische termen, zullen leken er maar weinig van [47] begrijpen.⁸ De kwantitatieve meting door middel van de FDRE is dus wel objectief en transparant, maar tegelijkertijd tamelijk beperkt. De FDRE vangt maar een klein deel van het construct ‘begrijpelijkheid’.⁹

Toch hebben ‘quick and dirty’ modellen zoals die van de FDRE ook hun charmes. Als beleidsmakers bijvoorbeeld willen bereiken dat – laten we zeggen – juridische teksten zoals algemene voorwaarden een bepaalde mate van begrijpelijkheid hebben, zouden ze er voor kunnen kiezen om te eisen dat

⁷ De formule luidt: $206,83 - (0,93 * \text{gemiddelde zinslengte}) - (77 * \text{gemiddeld aantal lettergrepen per woord})$.

⁸ Vgl. R. Flesch, 'A new readability yardstick', *Journal of Applied Psychology* 1948, p. 221 e.v.; M.E.J. Masson en M.A. Waldron, 'Comprehension of Legal Contracts by Non-Experts: Effectiveness of Plain Language Redrafting', *Applied Cognitive Psychology* 1994, p. 67 e.v.

⁹ Zie bijv. A. Davison en R.N. Kantor, 'On the Failure of Readability Formulas to Define Readable Texts: A Case Study from Adaptations', *Reading Research Quarterly* 1982, p. 187 e.v.; T. Oakland en H. Lane, 'Language, reading, and readability formulas: Implications for developing and adapting tests', *International Journal of Testing* 2004, p. 239 e.v. ; L.J. Sirico Jr., 'Readability Studies: How Technocentrism Can Compromise Research and Legal Determinations', *Quinnipiac L. Rev.* 2007, p. 26 e.v.

een tekst een minimumscore heeft op de FDRE-schaal.¹⁰ Het is namelijk eenvoudiger om vast te stellen of aan die numerieke eis is voldaan dan aan de eis 'de tekst moet op B2 niveau van het ERK geschreven zijn'. De eerste eis kan men met een eenvoudig computerprogramma toetsen. Om te zien of aan die laatste eis is voldaan, moet een taaldeskundige een 'diagnose' stellen. Is dan voldoende dat taalbureau X een verklaring afgeeft? Of moet taalbureau X eigenlijk gecertificeerd zijn om dergelijke verklaringen af te leggen? En wat als bij blinde hertoetsing een B1 of C1 uitslag volgt?

2.2. Taalbeheersing

Naast tekstniveau is er taalbeheersing, en die betreft de capaciteiten van lezers. Het taalbeheersingsniveau van 'gewone mensen' is niet bijzonder hoog. De recente OESO-rapportage *International Adult Literacy Survey (IALS)* laat een internationale vergelijkende methode zien, waarbij geletterdheid van individuen op een ladder [48] van vijf niveaus geplaatst kan worden. Niveau vijf is het hoogste niveau, en om standaardcontracten te begrijpen is niveau vier vereist.¹¹ Volgens Engels onderzoek heeft *minder dan 20 procent* van de Engelse bevolking dat vereiste taalbeheersingsniveau.¹² Vermoedelijk zijn de resultaten voor ons land wel iets beter, omdat geletterdheid in ons land beter is dan in Engeland.¹³ Toch laat ook onderzoek in ons land zien dat juridische documenten door veel mensen niet begrepen worden en de structuur niet eenvoudig doorzien wordt.¹⁴ Vermoedelijk is hier het achterliggende probleem dat juristen moeilijke teksten schrijven en niet altijd de moeite nemen om het beheersingsniveau van het lezerspubliek in acht te nemen (ik kom daar nog op terug).¹⁵ Ik denk eigenlijk dat er in de opleiding van juristen ook weinig aandacht besteed wordt aan 'wendbaarheid in schrijfvaardigheid'. Persoonlijk ben ik al blij als ik studenten zo ver heb dat redelijk foutloos wordt geschreven; aandacht in de opleiding voor tekstniveau is beperkt.

Idealiter sluit het tekstniveau aan bij het taalbeheersingsniveau van het lezerspubliek. Voor zover het onleesbare juridische teksten betreft die voor andere gespecialiseerde juristen bedoeld zijn, is dat niet erg problematisch. Dat geldt ook voor contracten tussen professionele partijen met juridische bijstand. Maar juridische teksten die gebruikt worden tegenover consumenten mogen wat dit betreft wel eens onder de loep worden gelegd. Dat is precies wat ik heb gedaan.

[49]

¹⁰ Zo kennen in de Verenigde Staten sommige staten de verplichting voor financiële dienstverleners om hun voorwaarden zo op te stellen dat zij ten minste een bepaalde FRE score (de Engelse variant van de FDRE) behalen. Zie bijv. de regels in de staat Arkansas: Ark. Code Ann. § 23-80-206.

¹¹ *OECD Skills Outlook 2013 - First Results from the Survey of Adult Skills*, OECD Publishing 2013.

¹² Office of Fair Trading (OFT), *Consumer Contracts (OFT1312)*, London 2011, p. 30.

¹³ *OECD Skills Outlook 2013 - First Results from the Survey of Adult Skills*, OECD Publishing 2013. Vgl. D. Fouarge e.a., *Laaggeletterdheid in Nederland*, Utrecht / Maastricht: ECBO / ROA Maastricht University 2011, p. 7.

¹⁴ H. Pander Maat, 'Naar gebruiksvriendelijke hypotheekinformatie - vier soorten praktijkstudies', *Tijdschrift voor Taalbeheersing* 2012, p. 74 e.v.

¹⁵ Vgl. L. Lentz, *Let op: Begrip verplicht! - Begrijpelijkheid als norm in de wet (oratie Utrecht)*, Utrecht: 2011, p. 30.

2.3. Een voorbeeld: algemene voorwaarden

Als we ervan uitgaan dat consumenten algemene voorwaarden daadwerkelijk lezen – al is het maar zodra ze in conflict komen met hun wederpartij – dan is het relevant of zij in staat zijn deze voorwaarden te begrijpen.¹⁶ In dat licht is het dienstig om een blik te werpen op de FDRE score van algemene voorwaarden in consumentencontracten. Ik heb daartoe 31 sets algemene voorwaarden uit de periode 1985-2012 geanalyseerd. Het betreft 9 sets uit de bankwereld, 12 uit de nutssector en 10 uit de reiswereld. De sets werden uit (gescande) pdf omgezet in kale tekstverwerkingsbestanden en voorbereid voor invoering in een online ‘readability’ applicatie.¹⁷ De uitkomsten waren toch licht ontmoedigend: de FDRE waarden bij deze 31 sets varieerde tussen 22,04 en 43,85 (M=31,05; SD=4,54). Dat is dus een gemiddelde van 31 op een schaal van 120, en dat is niet geweldig.¹⁸ Het komt in de buurt van de magere score van Van Maanen in het bovengenoemde citaat en het zal ongeveer overeen komen met tekstniveau C1-C2. Voor talige juristen begrijpelijk, maar voor veel particuliere klanten van nutsbedrijven, banken en reisorganisaties **[50]** te hoog gegrepen. Dit lijkt het beeld te bevestigen dat juristen vooral schrijven voor juristen, dat wil zeggen om waterdichte tekst te creëren.

Dat sluit aan bij een ander punt dat ik hiervoor al aanstipte, namelijk dat juristen in hun eigen taalwereld leven. Uit onderzoek lijkt te volgen dat het taalbeheersingsniveau van juristen niet alleen beter en verfijnder is dan dat van lager en anders opgeleiden, maar ook dat juristen niet snel een verschil in tekstniveau als zodanig *ervaren*.¹⁹ Dit laatste strookt in elk geval met mijn eigen ervaringen. Onlangs legden wij in een experiment met blinde toewijzing bij wijze van proef aan vierdejaars studenten privaatrecht twee sets algemene verzekeringsvoorwaarden voor, waarbij de ene set een tekstniveau (versie ‘makkelijk’) op B1 had en de andere (versie ‘moeilijk’) op C1 niveau. Vervolgens vroegen we de studenten om een aantal vragen te beantwoorden over tekstniveau. We

¹⁶ Over de vraag of consumenten de algemene voorwaarden lezen voorafgaand aan het maken van hun contractkeuze, is veel (ontluisterend) bewijsmateriaal voorhanden. Zie bijv. A.J. Rijsterborgh, 'De terhandstellingsplicht van algemene voorwaarden bij consumentenovereenkomsten', in: W.H. Van Boom e.a. (red.), *Boobytraps, valkuilen en instinkers in het burgerlijk recht*, Den Haag: BJu 2012, p. 71 e.v.; S.I. Becher en E. Unger-Aviram, 'The Law of Standard Form Contracts: Misguided Intuitions and Suggestions for Reconstruction', *DePaul Business & Commercial Law Journal* 2010, p. 199 e.v.; R.A. Hillman, 'On-line Consumer Standard-Form Contracting Practices: A Survey and Discussion of Legal Implications', in: J.K. Winn (red.), *Consumer protection in the age of the 'Information Economy'*, Aldershot: Ashgate 2006, p. 283 e.v.; H.A. Luth, *Behavioural Economics in Consumer Policy - The Economic Analysis of Standard Terms in Consumer Contracts Revisited (diss. Rotterdam)*, Antwerpen: Intersentia 2010, par. 5.5; M.G. Faure en H.A. Luth, 'Behavioural Economics in Unfair Contract Terms', *Journal of Consumer Policy* 2011, p. 337 e.v.

¹⁷ Deze online tool van de Universiteit Gent is te vinden op <http://lt3.hogent.be/en/demos/readability/>. De metingen werden verricht met tekst waaruit artikelnummers en koppen waren weggehaald om de meting zo betrouwbaar mogelijk te maken. Kleine afwijkingen blijven niettemin denkbaar, bijv. daar waar de tekst toch nog onbedoeld harde returns bevatte.

¹⁸ Bij het bestuderen van de 31 sets algemene voorwaarden viel mijn oog overigens op een afnemende FDRE score in tijd gemeten, maar die is niet significant. Een bivariate correlatietoets van jaartal en FDRE gaf $r(29)=-,29, p=.10$.

¹⁹ M. Malsch e.a., 'Hoe helder is de rechter? Leesbaarheid en begrijpelijkheid van strafrechtelijke vonnissen', *NJB* 2004, p. 1112 e.v. Vgl. in eenzelfde richting J. Kamalski e.a., 'Coherentiemarkering in informerende en persuasieve teksten', *Tijdschrift Voor Taalbeheersing* 2004, p. 85 e.v.

konden geen significant verschil tussen gemiddelden vinden. Wil men de werking van vereenvoudigd tekstniveau onderzoeken, dan kan men beter geen juristen als proefpersoon nemen, is mijn voorzichtige conclusie. Toen we dezelfde tekst namelijk aan twee groepen juridische leken voorlegden, was er *wel* significant verschil in ervaren moeilijkheidsgraad.²⁰

[51]

3. Van Maanen vs Hoge Raad

In het voorgaande heb ik een retorische truc uitgehaald om de schijn te wekken dat de geschriften van Van Maanen onleesbaar zijn: door één waarneming te isoleren en die centraal te stellen in mijn betoog, zou de onoplettende lezer kunnen denken dat Van Maanen al jaren een bedroevende FDRE-score van 26 uit 120 behaalt. Maar één enkele waarneming is niet per definitie een goede weergave van het gemiddelde van de gehele ‘populatie’ van zijn publicaties. Dus één citaat voor waar aannemen is net zo gevaarlijk als wanneer men het arrest *Lindenbaum/Cohen* zou bestuderen en vervolgens zou concluderen dat dit arrest baanbrekend is.²¹ Dat is een uitspraak over een hele populatie uitspraken op basis van $N=1$, en dat lijkt me gevaarlijk. Als bijvoorbeeld blijkt dat het betreffende arrest geen navolging kreeg en dat de Hoge Raad vervolgens weer is ‘terug omgegaan’, dan is het arrest per saldo niet baanbrekend. Hetzelfde zou gelden als blijkt dat de ‘maatschappelijke onzorgvuldigheid’ ook voordien reeds en nadien in ‘s Hogen Raads rechtspraak als maatstaf voor onrechtmatigheid werd gehanteerd. Een ‘ $N=1$ verhaal’ is dan dus meer een anekdotische dan frequentistische waarneming van de werkelijkheid.

Daarom heb ik geprobeerd om een frequentistisch beeld van de FDRE-scores van Van Maanen te creëren, om deze vervolgens af te zetten tegen de scores van arresten van de Hoge Raad. Dat was niet eenvoudig. Om de teksten van Van Maanen in de eerder genoemde online ‘readability’ applicatie in te voeren, was allereerst nodig dat de hand werd gelegd op een publiekelijk toegankelijke publicatielijst van Van Maanen en op de bijbehorende tekstbestanden in OCR-formaat (Optical Character Recognition). Ziehier het lot van eenieder die bronnenonderzoek wil doen – of het nu rechtshistorici zijn of civilologen, ze kampen allemaal met de gebrekkige toegankelijkheid van bronnen... Uiteindelijk is het [52] mij gelukt om een steekproef²² te nemen van ongeveer 3

²⁰ Bij 59 proefpersonen (vierdejaars rechtenstudenten; internetexperiment Erasmus Universiteit najaar 2012): ANOVA leverde $F(1,57) = 2,395$, $p = ,127$ (tweezijdig). Onze hypothese was dat studenten een verschil in tekstniveau zouden ondervinden, maar de bevindingen weerspreken die hypothese ($p = .06$; eenzijdig getoetst). In het kader van een ander onderzoek hebben we in 2013 aan ruim 1500 personen exact dezelfde twee sets voorwaarden voorgelegd. Als we binnen die groep alleen naar personen kijken met MBO en lager opleidingsniveau ($n=928$), dan volgt dat het verschil in gemiddelden wel significant is ($F(1, 926)=8,02$, $p=,005$); eenzijdig getoetst dus $p=,002$. Hier past uiteraard de kanttekening dat de groepsgrootte in het tweede geval veel groter is.

²¹ HR 31 januari 1919, NJ 1919/161, W. 10365 (en W. 10368) (*Lindenbaum/Cohen*).

²² De eerlijkheid gebiedt mij te zeggen dat ik niet weet of die steekproef groot genoeg is, want ten tijde van het schrijven deze bijdrage wist ik niet hoeveel publicaties Van Maanen in totaal op zijn naam heeft staan. De publicatielijst op de medewerkerspagina van de Maastricht University gaat niet verder terug dan tot 2001. *Leading in Learning* is de Universiteit Maastricht misschien wel, maar zeker niet in *Open Access* en *Archiving*.

publicaties per jaar. De steekproef was aselekt: ik kreeg de beschikking over bestanden in OCR-formaat, gerangschikt op jaar, daaruit nam ik steeds op goed geluk 3 publicaties, waarna ik zonder nadenken een alinea uit de tekst kopieerde en in de online 'readability' applicatie invoerde. Er zijn ongetwijfeld beter geprotocolleerde methoden van steekproefneming, maar voor mijn gevoel volstond dit voor het doel van deze bijdrage.

Aldus kreeg ik de beschikking over 49 alinea's tekst (waarnemingen) uit de periode tussen 1986 en 2013. Bij wijze van voorbeeld laat ik hier twee waarnemingen zien:

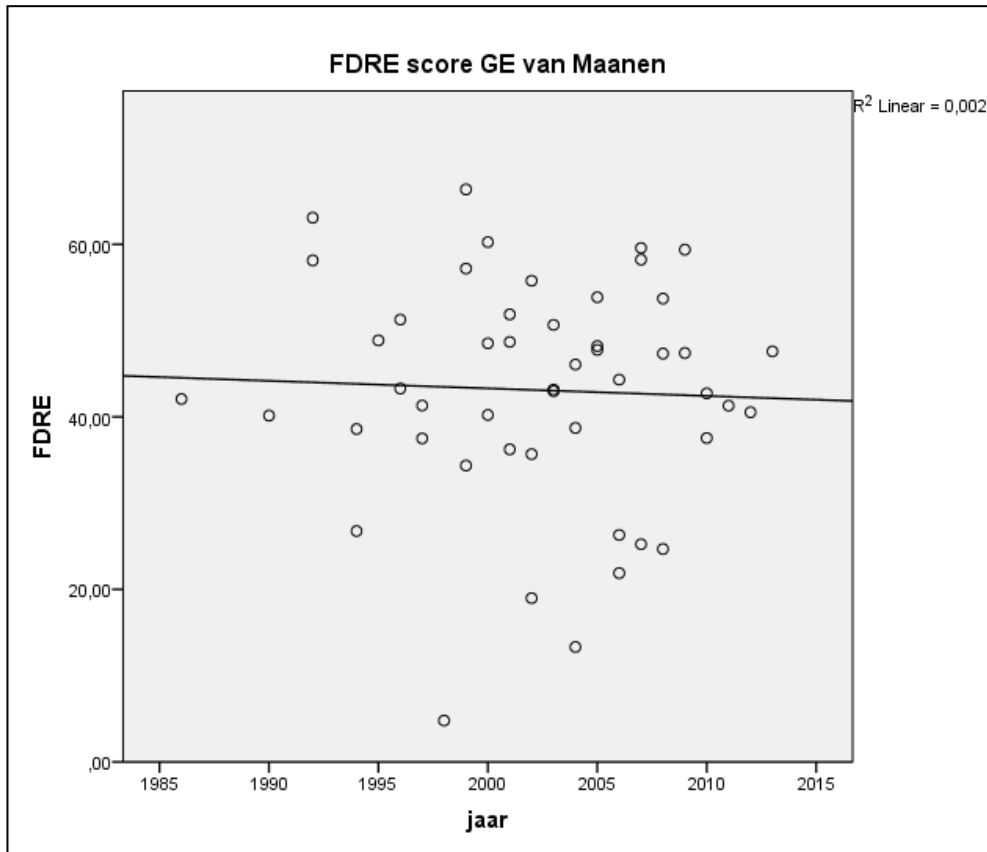
FDRE 4.79 (Van Maanen, NTBR 1998, p. 150; annotatie)	FDRE 66.36 (Van Maanen, NTBR 1999, p. 137; naschrift)
Dat scheidt niet alleen problemen met betrekking tot de afstemming van verschillende uitspraken tot elkaar - Spier wees er bijvoorbeeld op dat bij verkeersaansprakelijkheid, latere uitspraken soms niet verenigbaar waren met eerdere beslissingen - maar ook ontstaan er problemen met de aansprakelijkheidsverzekering. Met terugwerkende kracht kunnen immers bepaalde schadeveroorzakende gebeurtenissen onrechtmatig worden beoordeeld en tot vergoeding van de aansprakelijkheidsverzekeraar nopen. Vooral in zaken van milieuschade en beroepsaansprakelijkheid kan dit tot grote problemen leiden.	Het tweede argument is dat het in het vermogensrecht uiteindelijk draait om de poen en dat er daarom in het vermogensrecht geen plaats is voor het belang om de schuld vastgesteld te willen zien van de dood van een kind. Maar dat is een wei heel doorzichtige redeneertruc van mijn geachte opponente. Dat is natuurlijk per definitie waar! Maar dat is ook niet wat ik gesteld heb; ik heb gesteld dat het in het privaatrecht om meer gaat dan alleen vermogensrechtelijke belangen. Dit onder verwijzing naar diverse plaatsen waar ook niet-vermogensrechtelijke belangen in het wetboek beschermd worden. Tenslotte maak ik gebruik van de gelegenheid om een correctie aan te brengen. Ik schreef dat het ging om een dertienjarig kind, maar in werkelijkheid betrof het een driejarig kind. Aan de strekking van mijn betoog doet dat evenwel niet af.

[53] Zoals men kan zien, varieert de FDRE nogal. De betreffende teksten zijn dan ook de buitenwaarden, zoals blijkt uit de beschrijvende statistieken (figuur 1):

Descriptive Statistics						
	N	Range	Minimum	Maximum	Mean	Std. Deviation
FDRE	49	61,57	4,79	66,36	43,1092	13,01372

Figuur 1 - FDRE waarden Van Maanen

De gemiddelde FDRE-score van Van Maanen is dus 43. In een scatterplot (figuur 2) ziet dat er als volgt uit (waarbij de vrijwel horizontale lijn de kleinste-kwadratenmethode van de waarnemingen weergeeft):



Figuur 2 Scatterplot FDRE waarden Van Maanen

Maar wat zegt zo'n waarde van 43 nu eigenlijk? Laten we eerst maar eens proberen te vergelijken met mijzelf: hoe begrijpelijk is Van Boom eigenlijk? Om dat te meten, hoeft men aanmerkelijk **[54]** minder moeite te doen dan bij Van Maanen. Vrijwel alle publicaties van Van Boom zijn namelijk in OCR-formaat op het internet (om niet!) te verkrijgen.²³ Van 2 à 3 aselect gekozen publicaties per jaar uit de periode 1993-2010 analyseerde ik de gehele tekst. Zo kwam ik tot 46 waarnemingen en een gemiddelde FDRE van ongeveer 42,9 (zie figuur 3). Dat is dus *iets slechter* dan Van Maanen.²⁴

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
FDRE	46	33,03	56,09	42,9215	5,86671

Figuur 3 FDRE waarden Van Boom

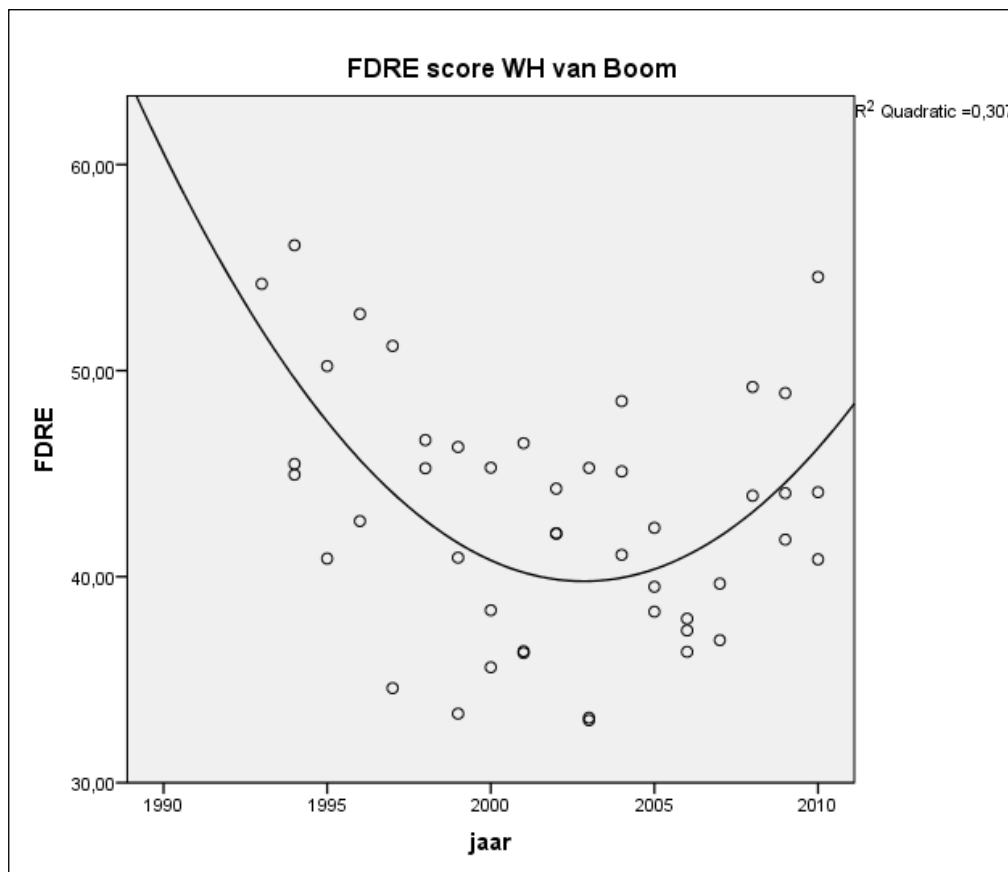
Doordat meer tekst per publicatie is ingevoerd, wordt de meting nauwkeuriger en – aannemende dat een auteur een bepaalde stabiele schrijfstijl heeft – minder variabel. Dit verklaart waarschijnlijk dat bij de meting van Van Boom de standaarddeviatie minder groot is dan bij Van Maanen. Maar het

²³ Bijv. via www.professorvanboom.eu.

²⁴ Het goede nieuws is bovendien dat het verschil tussen Van Maanen (M=43,10; SD=13,01) en Van Boom (M=42,92; SD=5,86) niet significant is ($t(67,6) = -0,90, p = ,927$). Dus eigenlijk ben ik net zo goed als Gerrit, althans ik heb niet kunnen vaststellen dat hij beter is.

verklaart niet waarom Van Boom in de loop der jaren slechter lijkt te zijn gaan schrijven. Op het oog (zie scatterplot; figuur 4) ben ik namelijk minder begrijpelijk geworden, kennelijk met als dieptepunt 2002 (het jaar van mijn benoeming tot hoogleraar).... Gelukkig hoef ik daar geen verklaring voor te zoeken omdat het hier niet om een significante bevinding gaat.²⁵ *Gered door statistieken*, zullen we maar zeggen.

[55]



Figuur 4 scatterplot FDRE waarden Van Boom

En dan de hamvraag: hoe doet de Hoge Raad het? Hier ben ik goed voor gaan zitten. Uit een periode van 20 jaar (1990-2010) analyseerde ik 687 arresten en beschikkingen van de civiele kamer van de Hoge Raad. De zaken werden aselekt gekozen, zij het dat steeds intervallen van 5 jaar werden genomen en de art. 81 RO zaken buiten beschouwing werden gelaten.²⁶ Het resultaat is een stabiel gemiddelde van 50,8 op de FDRE-schaal (figuur 5; figuur 6).

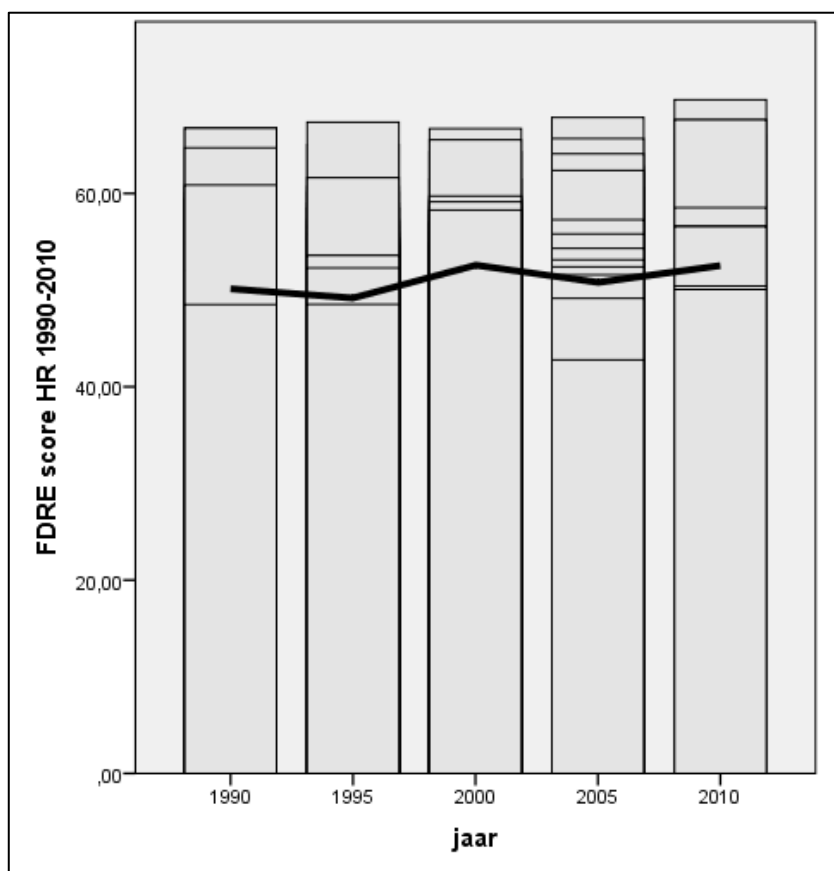
²⁵ Een lineaire bivariate correlatietoets van jaartal en FDRE gaf geen significante uitkomsten. Hoewel op het oog een curvilineair verband ook tot de mogelijkheden behoort (zie de curvelijn in de scatterplot), gaf een gekwadateerde curve estimation evenmin significante uitkomsten.

²⁶ Op basis van een grof geschat aantal uitspraken (rol+ rekest - 81 RO=) van 6000 a 7000, zou een steekproef van ongeveer 370-400 volstaan. Ik nam uit de jaren 1990, 1995, 2000, 2005 en 2010 steeds een steekproef van ± 100-150 arresten.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
FDRE	687	31,02	69,69	50,846	5,82746

Figuur 5 FDRE waarden Hoge Raad (civiel)

[56]

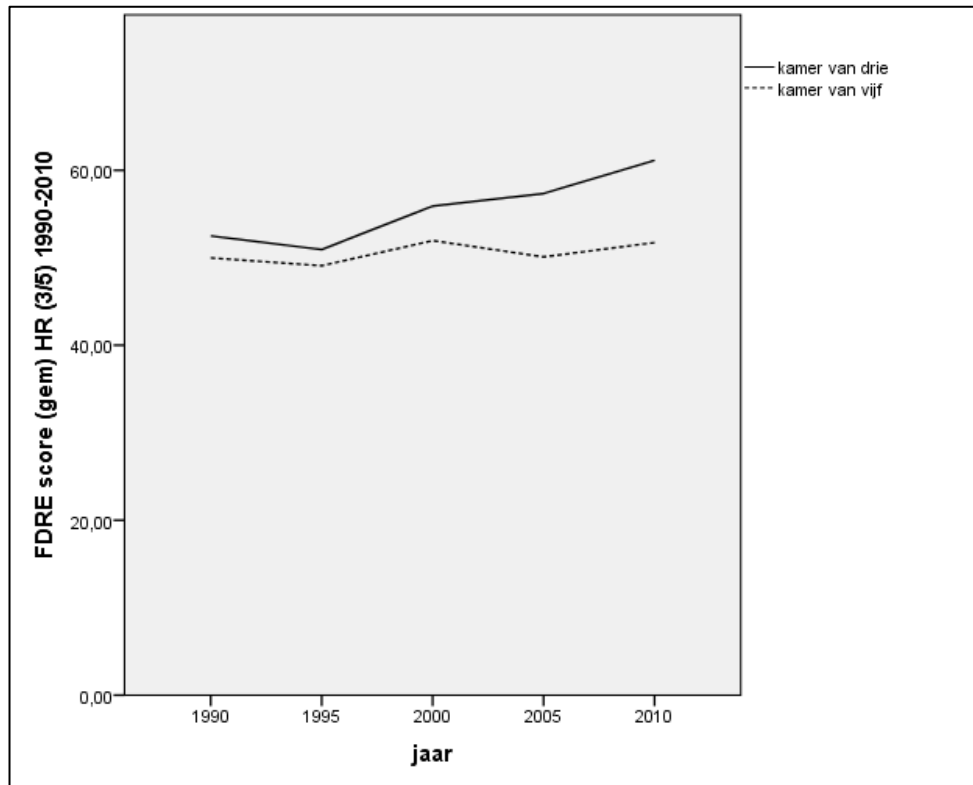


Figuur 6 Histogram FDRE waarden Hoge Raad (civiel) 1990-2010

Ik heb bovendien vastgesteld dat er verschil tussen kamers van 3 en 5 bestaat (zie figuur 7). Er is significant verschil tussen de FDRE score van de kamers van 3 ($M=56,06$, $SD=6,8$) en de kamers van 5 ($M=50,4$, $SD=5,5$); $t(685)=7,31$, $p<0,01$.²⁷

²⁷ Hier bestaat overigens wel een complicatie wat betreft de data: de verhouding in onze steekproef tussen kamers van 3 en 5 was 57:630 en dat wijkt ernstig af van de werkelijke verhouding. Die was namelijk in 2000: 169/309; 2005: 123/329; 2010: 291/291 (bron: Jaarverslag HR 2002; 2005; 2009-2010). Een goede verklaring

[57]



Figuur 7 FDRE score HR kamers 3 vs 5

Om de resultaten kort samen te vatten: de FDRE-waarden van de Hoge Raad zijn al jaren stabiel én ze zijn significant beter dan Van Maanen (en dus ook begrijpelijker dan Van Boom).²⁸ En om het nog wat preciezer te zeggen: de Hoge Raad is zo'n 16 tot 30% begrijpelijker dan Van Maanen...

4. Relativering en reflectie

Het voorgaande moet natuurlijk serieus gerelativeerd worden. Misschien moet het zelfs in het geheel niet serieus genomen worden. Niettemin is er een serieuze boodschap bij het ietwat [58] frivole

voor dit verschil zal moeten worden gezocht in het gegeven dat uitspraken van kamers van 3 relatief vaker afgedaan worden met 81 RO of (voor de oudere zaken die we uit de NJ hebben gehaald:) relatief minder snel gepubliceerd worden. Aangezien art. 81 RO –zaken juist uit onze steekproef weggelaten werden, zijn de kamers van 5 als het ware oververtegenwoordigd. Om vast te stellen of dit werkelijk een probleem is, heb ik de 57 zaken van de kamer van drie vergeleken met een aselechte steekproef van 57 zaken van de kamer van vijf, en de verschillen in gemiddelde bleken ook toen significant: kamers van 3 ($M=56,06$, $SD=6,8$) en kamers van 5 ($M=49,78$, $SD=4,9$); $t(112)=5,61$, $p<0.01$.

²⁸ Een ANOVA van de 592 waarnemingen van de Hoge Raad, de waarnemingen op Van Maanen en Van Boom levert $F(2, 864) = 53,746$, $p<,001$.

onderzoekje dat ik hiervoor presenteerde. De werkelijkheid is altijd complex en als men deze met objectieve precisie wil waarnemen, dan is modelleren nodig: het versimpelen, het weglaten van ruis, maar ook het negeren van nuance en het zoeken van meetbaarheid van wat zich niet goed laat meten. Zo kan men uiteraard de vraag stellen of het meten van FDRE voldoende inhoudsvaliditeit heeft: krijgt men van het construct 'begrijpelijkheid' werkelijk een goed beeld door een rekenkundige weging van de gemiddelde zinslengte en het gemiddeld aantal lettergrepen per woord uit te voeren?²⁹ Nou, in elk geval geen *volledig* beeld. Misschien is verdedigbaar dat de constructvaliditeit van de FDRE-score wel aanvaardbaar is als men juridische teksten onderling vergelijkt – als we aannemen dat het gebruik van technisch vocabulaire niet wezenlijk verschilt tussen de teksten van Van Maanen, Van Boom en de Hoge Raad, dan moet het verschil in FDRE-score wel gelegen zijn in zinsbouw en woordkeuze.

Een tweede kanttekening is deze. Het doen van kwantitatief onderzoek zonder theorie over wat men denkt te gaan vinden en waarom men dat denkt, heeft beperkingen. En soms is het ronduit gevaarlijk. Bij het tellen en turven van de rechtspraak van de Hoge Raad had ik geen uitgesproken ideeën over de vraag of er verschil in FDRE-score tussen kamers van 3 en kamers van 5 gevonden zou worden. En nu ik dat verschil wel heb gevonden, is het eigenlijk nodig om een verklarende hypothese te formuleren en die vervolgens te toetsen. Wat kan het verschil verklaren? Is het misschien omdat kamers van 3 vooral eenvoudige zaken behandelen, die tot weinig discussie in raadkamer en dus tot weinig amendering van concepten aanleiding geven? Als dat de hypothese is, moet men ook de daarin besloten liggende aanname toetsen dat amendering van concepten tot langere zinnen (en / of langere woorden) leidt.

[59] Een verdere reflectie bij het voorgaande moet zijn dat het doen van kwantitatief onderzoek binnen het juridisch discours tijdrovend is en niet snel tot tastbare en relevante resultaten aanleiding zal geven. Willen rechtenfaculteiten echt ernst maken met dergelijk onderzoek, dan moeten wetenschappers betere prikkels krijgen om *minder en met meer impact* te schrijven.

Een laatste reflectie betreft het nut van dit type onderzoek. Wat hebben we hier nu aan? Het doen van civilologisch onderzoek vergt veel oefening, en ook als het nut voor beantwoording van De Grote Onderzoeksvragen niet direct evident is, moet die oefening toch plaatsvinden. Dus deze bijdrage heeft niet alleen de score Van Maanen (43), Van Boom (42,9) en Hoge Raad in 3-formatie (56) opgeleverd. De bijdrage heeft als het goed is, ook laten zien hoezeer onderzoek naar de rechtswerkelijkheid door beschikbaarheid van data wordt beperkt, hoezeer de keuze voor de opzet van meting, codering en steekproefneming daardoor wordt beïnvloed en hoezeer men al doende het zicht op de werkelijkheid kan verliezen wanneer men constructvaliditeit niet serieus neemt. Want laten we wel wezen, wat zegt FDRE nu over écht begrijpelijke wetenschap? (sprak de verliezer...)³⁰

²⁹ Over constructvaliditeit bijv. M. Leary, *Introduction to Behavioral Research Methods*, Boston: Pearson 2012, p. 62-63; C.C.J.H. Bijleveld, *Methoden en Technieken van Onderzoek in de Criminologie*, Den Haag: BJU 2009, p. 41-42.

³⁰ Hoewel.... deze bijdrage scoort ruim 53 op de FDRE-schaal! Zit ik nog steeds in de lift?