De culturele verankering van evaluatie

Nederlandse en Vlaamse dagbladrecensies in computationeel perspectief

Joris Veerbeek

DE CULTURELE VERANKERING VAN EVALUATIE

NEDERLANDSE EN VLAAMSE DAGBLADRECENSIES IN COMPUTATIONEEL PERSPECTIEF

Joris Veerbeek

Scriptie RMA Nederlandse Literatuur en Cultuur
Universiteit Utrecht
Begeleider: Geert Buelens

ABSTRACT

INTRODUCTION While Flanders and The Netherlands share the same language and a large part of their history, the independence of Flanders/Belgium in 1830 has led both countries to develop largely independently – politically as well as culturally. Many scholars have theorized about the question whether modern-day Dutch and Flemish literature could nonetheless be considered one of the same. However, little is known about the actual differences and similarities between both literary fields, since structural comparisons are lacking.

OBJECTIVE I employ computational methods to compare post-war Dutch and Flemish literary criticsm on a large corpus containing 22.361 reviews.

METHODS I use and compare both state-of-the-art and well-established Natural Language Processing techniques (most notably: Word2vec, Bi-GRU and BERT) to automatically classify and categorize evaluative statements in literary reviews based on a classification scheme developed by literary scholars and communication researchers. This classification scheme categorizes both the *aspects* critics use when evaluating literary prose, such as the style, the author or the characters, as well as the *characteristics* that are assigned to those aspects, e.g. humoristic, emotional or realistic. The employed machine learning models achieved a relatively high accuracy, allowing us to classify and compare which aspects and characteristics Flemish and Dutch critics take into account.

RESULTS The comparision reveals that Flemish reviewers are significantly more concerned with the ideological (e.g. the religious, political, moral and didactic) characteristics of literary works than Dutch reviewers. When discussing these political characteristics, Dutch critics tend to highlight the relationship of the work to current events, whereas Flemish critics speak more about (political) engagement. Furthermore, the diachronic analysis shows that extra-literary evaluations have dominated the nature of Flemish literary reviews longer. In Dutch reviews, extra-literary evaluations have steadily decreased since 1946, subsequently allowing for the rise of more autonimistic evaluations (e.g. *structure*, *style*). In Flemish reviews, this is only the case from the second half of the sixties.

WHY IT MATTERS The results show how the 'variable of nation' plays a role in literary criticsm. They give insight into the extent of which reception and evaluation is embedded within a cultural-societal framework.

INHOUDSOPGAVE

1 INLEIDING 1

I ACHTERGROND

2 LITERATUURKRITIEK VANUIT KWANTITATIEF PERSPECTIEF 10
2.1 Evualatieve uitspraken 11
2.2 Aspecten en eigenschappen 12
2.3 Literatuuropvattingen 13
3 DIGITAL HUMANITIES 16
3.1 Distant reading 17
3.2 Recente kritiek 19
3.3 Pragmatisme en plaatsbepaling 20

II METHODE

4 DE CLASSIFICATIE VAN RECENSIES 24
4.1 Dataverzameling 25
4.2 Probleembeschrijving 28
4.3 Features 32
4.4 Modellen 40
4.5 Evaluatie 45
4.6 Resultaten 47
5 CORPUS 53
5.1 Dataverzameling 54
5.2 Corpusstatistieken 60
6 ANALYSEMETHODE 63
6.1 Analyseniveau 63
6.2 Statistische toetsen 65
6.3 Effectgroottes 65

III RESULTATEN

7 ASPECTEN IN NEDERLANDSE EN VLAAMSE RECENSIES 70
7.1 Synchrone analyse 72
7.2 Diachrone analyse 73
7.3 Verschillen tussen Vlaamse en Nederlands schrijvers 79
8 EIGENSCHAPPEN IN NEDERLANDSE EN VLAAMSE RECENSIES 82
8.1 Synchrone analyse 83
8.2 Diachrone analyse 88
8.3 Verschillen tussen Nederlandse en Vlaamse schrijvers 92

IV CONCLUSIE
9 CONCLUSIE 96

BIBLIOGRAFIE 105
A BIJLAGE 114
A.1 Eigenschapscodes per krant 114
A.2 Top veertig meest voorkomende woorden in politieke evaluaties 115
A.3 Distributies van waarschijnlijkheden 116

Hoofdstuk 1

INLEIDING

Wanneer een Nederlandse literatuurcriticus een boek – een verhalenbundel, een roman, een poëziebundel – beoordeelt, in hoeverre doet hij dat dan ook als een Nederlander? Uiteraard zal hij tijdens het beoordelen geen oranjekleurig tenue dragen of zachtjes het Wilhelmus neuriën. Dit "nationalisme"– als je het al zo kan noemen – speelt zich veel subtieler af: in onbewuste voorkeuren en zwaartepunten; in een gedeeld cultureel, historisch en politiek referentiekader met zeventien miljoen andere (al dan niet ingebeelde) Nederlanders. Een criticus, zo stelt literatuursociologisch onderzoek, opereert immers niet volledig autonoom, maar is als speler in het literaire veld ingebed in een breder cultureel-maatschappelijk veld, en zijn literaire oordelen en normen worden gevormd in interactie met deze verschillende contexten.[1] Maar betekent dit ook dat Nederlandse literatuurcritici boeken anders beoordelen en evalueren dan hun Vlaamse collega's?

Als we de promotie rondom de *Frankfurter Buchmesse* van 2016, waar Nederland en Vlaanderen voor de tweede keer in de geschiedenis gezamenlijk optraden als gastland, mogen geloven, is dit allerminst het geval. *Dit is wat we delen*, zo luidde het motto van de boekenbeurs. En wat delen ze? Niet het minste: 'Een taal, een geschiedenis, en een rijke literatuur en cultuur'. Hoewel er vrijwel altijd wat op dergelijke – ongetwijfeld goedbedoelde – PR-slogans valt af te dingen, is de overtuigdheid waarmee Nederland en Vlaanderen zich als een cultureel collectief presenteerden desalniettemin illustratief voor hoe er over de literaire en culturele verhoudingen tussen beide landen gedacht wordt. Dat de Engelse, Amerikaanse en de Ierse literatuur, aldus Kris Humbeeck, aparte eenheden vormen, 'zullen maar weinig mensen tegenspreken'.[2] Hetzelfde geldt voor de Duitse, Zwitserse en Oosten-

1 Esther Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse" (proefschrift, Radboud Universiteit Nijmegen, 2014), 15-16; Susanne Janssen, *In het licht van de kritiek: variaties en patronen in de aandacht van de literatuurkritiek voor auteurs en hun werken*, deel 14 (Uitgeverij Verloren, 1994); Gillis J Dorleijn en Cornelis Jacob van Rees, *De productie van literatuur. Het Nederlandse literaire veld 1800-2000* (Vantilt, 2006), 19.

2 Kris Humbeeck, "Een knuppel in het boekenhok: schrijf Vlaamse literatuur, geen Zuid-Nederlandse!", *De standaard*, 2004, https://www.standaard.be/cnt/dma21042005_021.

rijkse literatuur.[3] Maar ondanks wederzijdse culturele desinteresse[4] willen 'we (maar wie zijn die 'we'?) per se blijven spreken over die ene Nederlandse literatuur'.[5] En inderdaad: wie de geschiedenis van de *Frankfurter Buchmesse* bestudeert, ziet dat Nederland en Vlaanderen tweemaal gezamenlijk als gastland optraden, terwijl Oostenrijk, Zwitserland en Ierland zich allen als een individueel geheel – als een eigen literatuur – presenteerden.

Binnen de neerlandistiek is de literaire verhouding tussen Nederland en Vlaanderen desalniettemin al geruime tijd uitvoerig onderwerp van discussie. Niet voor niets duidt Geert Buelens de vraag of de Nederlandse en Vlaamse literatuur gezamenlijk besproken moeten worden aan als 'de Moeder van alle Historiografische Vragen in ons vakgebied'.[6] Ruwweg zijn er binnen dit debat twee historiografische posities van elkaar te onderscheiden. De eerste positie meent dat Nederland en Vlaanderen – historisch, cultureel, institutioneel – dusdanig van elkaar verschillen dat beide literaire systemen in hoge mate verschillend functioneren, en de Nederlandse en Vlaamse literatuur dan ook primair als 'aparte grootheden' beschouwd moeten worden.[7] De tweede erkent deze verschillen, maar meent dat deze verschillen zich juist profileren binnen een (historische) eenheid – hetgeen we, naar Bert Vanheste, kunnen omdopen tot het 'eenheid in verscheidenheid'-argument.[8] Waar de eerste positie veelal leidt tot het isoleren van de Vlaamse dan wel de Nederlandse literatuur, lijkt de tweede positie vrijwel onvermijdelijk te resulteren in een soms enigszins geforceerde synthese van twee losse verha-

3 Overigens kunnen we wel gelijkaardige sentimenten in beschouwingen over Oostenrijkse literatuur vinden. Vergelijk: 'the cultural neutralisation of Austrian literature in German university libraries is remarkable: Authors such as Franz Kafka, Robert Musil, Ödön von Horváth or Heimito von Doderer are fading behind categories of a very specific history of German literature such as Expressionism or *Neue Sachlichkeit*'. Geciteerd uit Wolfgang Müller-Funk, "Austrian Literature in a Trans-cultural Context", in *The architecture of modern culture: Towards a narrative cultural theory*, deel 3 (Walter de Gruyter, 2012), 248.

4 Humbeeck baseert zich in deze stelling op algemene indrukken, maar een meer cijfermatige onderbouwingen zijn gegegeven in Marc Verboord en Suzanne Janssen, "Informatieuitwisseling in het huidige Nederlandse en Vlaamse literaire veld. Mediagebruik en gelezen boeken door literaire lezers en bemiddelaars", in *Een of twee Nederlandse literaturen? Contacten tussen de Nederlandse en Vlaamse literatuur sinds 1830* (Peeters, 2008), 307–321; Carlo Van Baelen, *Begrensd literair grensverkeer. De zichtbaarheid van 150 literaire titels in Vlaanderen en Nederland.* (Taalunie, 2014).

5 Humbeeck, "Een knuppel in het boekenhok: schrijf Vlaamse literatuur, geen Zuid-Nederlandse!"

6 Geert Buelens, *Van Ostaijen tot heden: zijn invloed op de Vlaamse poëzie* (Vantilt, 2008), 16. In eenzelfde context – maar aanzienlijk minder positief geformuleerd – spreekt Hugo Brems over 'chronische debatten'. Zie: Hugo Brems, "Nederlandse literatuur in Vlaanderen?", *Literatuur* 93, nummer 1 (1993), 314.

7 E.g. Ton Anbeek, *Geschiedenis van de Nederlandse literatuur tussen 1885 en 1985* (De Arbeiderspers, 1990), 16; Buelens, *Van Ostaijen tot heden: zijn invloed op de Vlaamse poëzie*; Hugo Brems, "Het centrum als periferie als centrum", *Dietsche Warande en Belfort* 141, nummer 2 (1996): 211–218.

8 E.g. Bert Vanheste, "Van de eenheidsworst, verlos ons, Ton", *Dietsche Warande en Belfort* 141 (1996): 231; Anna Musschoot, "Anbeeks keuze: enkele kanttekeningen", *Dietsche Warande en Belfort* 141, nummer 2 (1996): 233–242; Maria A. Schenkeveld-Van der Dussen, *Nederlandse literatuur, een geschiedenis* (Uitgeverij Contact, 1998); Wiljan Van den Akker en Gillis Dorleijn, "Consensus en conflict. Over de literatuurgeschiedschrijving van Nederland en Vlaanderen", *Nederlandse letterkunde* 4, nummer 2 (1999): 191–210.

len, aan elkaar geregen door varianten op beginzinnen als "Rond dezelfde tijd in Vlaanderen..".[9]

Een derde optie, een (poëticale) *vergelijking* tussen het Vlaamse en het Nederlandse literaire systeem, is schaars.[10] Hoewel er wel enkele pogingen zijn gedaan om poeticale overeenkomsten en verschillen tussen Nederland en Vlaanderen te formuleren,[11] dienen dezen veelal een programmatisch doel. Dat wil zeggen: ze dienen vooral ter ondersteuning voor één van de bovengenoemde twee historiografische posities. Wanneer Ton Anbeek enkele voorbeelden noemt die laten blijken dat de Nederlandse en de Vlaamse literatuur historisch en poëticaal sterk verschillen, werpt Anne Marie Musschoot tegen: 'Naast de voorbeelden van Anbeek kunnen [...] ook tegenvoorbeelden worden genoemd die niet zozeer wijzen op verschillen als wel op behoefte aan samenwerking, op wisselwerking en dus: op verwevenheid, samenhang en eenheid'.[12] Voor een preciezer begrip van de literaire en poëticale verhoudingen tussen Nederland en Vlaanderen, van de wisselwerking tussen literatuuropvattingen en maatschappelijke context, moeten we een manier vinden waarop we deze systematisch kunnen vergelijken; waarmee we niet meer afhankelijk zijn van individuele voorbeelden, maar deze – in zijn geheel – kunnen aggregeren en kwantificeren.

In deze scriptie wordt een dergelijke poging ondernomen. Ze vergelijkt, op systematische, grootschalige en kwantitatieve wijze, evualatieve uitspraken in Vlaamse en Nederlandse literatuurkritiek. Literatuurcritici vormen – in literatuursociologische termen die inherent lijken aan de bestudering van literatuurkritiek – cruciale actoren in het literaire veld, en zijn 'medebepalend voor de manier waarop volgende werken en auteurs worden besproken of de literatuurgeschiedenis ingaan'.[13] Omdat zij zowel een sterke invloed hebben op de normen en waardeoordelen van het literaire veld, als er zelf ook door worden beïnvloed, kan het bestuderen en vergelijken van Nederlandse en Vlaamse literatuurkritiek ons – in extensie – veel onthullen over verschillen in achterliggende poëticale opvattingen van beide culturen in geheel, zoals Hugo Brems eerder heeft beargumenteerd.[14]

Om literatuurkritiek te kunnen kwantificeren bouwt deze scriptie voort op het op het categoriseringsmodel dat is ontwikkeld door Yvette Linders

9 Ton Anbeek, "Het Vlaamse verschil", *Dietsche Warande en Belfort* 2 (1996): 206; Hugo Brems, ""Rond dezelfde tijd in Vlaanderen..."", in *Een of twee Nederlandse literaturen? Contacten tussen de Nederlandse en Vlaamse literatuur sinds 1830* (Peeters, 2008), 25.

10 Zie ook Geert Buelens, "Naar een Vergelijkende Neerlandistiek", in *Colloquium Neerlandicum* (2018), https://www.platformleest.org/lezing/naar-een-vergelijkende-neerlandistiek.

11 Hugo Brems, "De relatie tussen de Nederlandse en de Vlaamse literatuur. De teller en de noemer", *Tijdschrift over Nederlandse Letterkunde* 5, nummer 1 (1988): 34–38; Ben Van Humbeeck en Floor Van Renssen, "Op voet van gelijkheid? De Noord-Zuidkwestie in de Nederlandse en de Vlaamse literaire kritiek, 1980-1995", in *Een of twee Nederlandse literaturen? Contacten tussen de Nederlandse en Vlaamse literatuur sinds 1830* (Peeters, 2008), 263–290; Geert Buelens, "Gotspe of godsgeschenk? Enkele notities over zin en onzin van een Belgische literatuurgeschiedenis", in *Literaire belgitude littéraire. Bruggen en beelden. Vues du Nord* (Presses universitaires de Louvain, 2011), 67–80; Anbeek, "Het Vlaamse verschil".

12 Musschoot, "Anbeeks keuze: enkele kanttekeningen", 235.

13 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 14.

14 Brems, "De relatie tussen de Nederlandse en de Vlaamse literatuur. De teller en de noemer".

en Esther op de Beek.[15] In dit model worden evaluatieve uitspraken in recensies gecodeerd naar het aspect van de literaire tekst dat besproken wordt (de stijl, de symboliek, het plot), en naar de eigenschap die aan dat aspect wordt toegekend (helder, emotioneel, origineel). Waar Linders en Op de Beek hun corpus handmatig hebben gecodeerd, stelt deze scriptie zich methodologisch ten doel om deze categorisering automatisch te reproduceren. Hiervoor maak ik in deze scriptie dankbaar gebruik van het door Linders en Op de Beek gecodeerde recensiecorpus. Eerder ontwikkelde Peter Boot een regelgebaseerde methode om internetrecensies automatisch te annoteren volgens een variatie op het model van Linders en Op de Beek.[16] In deze scriptie wordt gekozen om, op basis van neurale taalmodellen en de eerdere annotaties, de computer het classificatiemodel te laten leren met recente technieken uit Natural Language Processing.

Daarmee analyseer ik vervolgens een groot recensiecorpus van Nederlandse en Vlaamse literaire recensies (1945-2018), bestaande uit 22.361 recensies. Daarbij beperk ik mij wel, net als Linders en Op de Beek, tot recensies die 1) Nederlandstalig proza bespreken, en 2) in kranten verschenen zijn. Hoewel ik mij bij Op de Beek aansluit dat 'de aard en functie van recensies sterk [is] gebonden aan het medium waarin recensies verschijnen', en prozarecensies inhoudelijk ook sterk verschillen van poëzierecensies,[17] is de reden daarvoor hoofdzakelijk technisch: de annotaties op basis waarvan de computer beslissingen maakt zijn ook beperkt tot dagbladrecensies van proza, waardoor het niet eenvoudig te veronderstellen is dat de computer evaluatieve uitspraken in tijdschriftrecensies (of: poëzierecensies) met dezelfde precisie kan classificeren.

De vergelijking tussen Nederlandse en Vlaamse literaire kritiek verloopt in deze scriptie zowel synchroon als diachroon: Hoe verschilt de Nederlandse met de Vlaamse literatuurkritiek door de tijd heen? Is er – naar Van Humbeeck en Van Renssen – sprake van toenemende convergentie in de literatuuropvattingen van Nederlandse en Vlaamse critici sinds de tweede helft van de jaren tachtig?[18] Zijn er bepaalde onveranderde eigenschappen en aspecten die bij Nederlandse literatuurcritici meer aandacht krijgen dan bij hun Vlaamse collega's, of andersom? Maar ook: hanteren Vlaamse critici voor Vlaamse auteurs andere normen dan voor Nederlandse auteurs, en vice versa? Waar Op de Beek aan de hand van het analysemodel analyseert hoe 'maatschappelijke veranderingen, poëticale verschuivingen en literaire ontwikkelingen de wijze van evalueren beïnvloeden' is de voornaamste onafhankelijke variabele van deze scriptie dus 'natie' (of, preciezer: herkomst

15 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse"; Yvette Linders, "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005" (proefschrift, Radboud Universiteit Nijmegen, 2014).

16 Peter Boot, "Distant reading Nederlandstalige online boekrespons", in *Achter de verhalen* (2018).

17 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 21 & 86.

18 Van Humbeeck en Van Renssen, "Op voet van gelijkheid? De Noord-Zuidkwestie in de Nederlandse en de Vlaamse literaire kritiek, 1980-1995"; Floor Van Renssen, *'Lezer, er zijn ook Belgen!': interactie tussen de Nederlandse en Vlaamse literatuur via literaire kritiek en uitgeverij (1980-1995)* (Maklu, 2013).

van de recensent.[19] Daarbij operationaliseer ik deze variabele aan de hand van de herkomst van recensent.[20]

In een internationale onderzoekscontext – en dan primair in het Angelsaksische taalgebied – hebben verschillende studies al op een gelijkaardige vergelijkende wijze gewezen op het belang van nationaliteit in de evaluatie van kunst in het algemeen, en literatuur in het bijzonder. Zo wordt de studie van Wendy Griswold in cultuursociologisch onderzoek geregeld in deze context aangehaald.[21] Aan de hand van een vergelijking van Amerikaanse, Britse en West-Indische recensies van het werk van één schrijver – de Barbadiaanse auteur George Lamming – tracht ze de sociaal-contextuele afhankelijkheid van betekenisvorming (en, daaraan verbonden, waardering) van culturele producten bloot te leggen. Zo blijkt dat Britse recensenten het evaluatieve zwaartepunt legden op de vormelijke, zogezegd 'literaire' eigenschappen van Lammings werk, terwijl hun Amerikaanse collega's zijn werk juist veel meer lazen langs de lijnen van raciale thematiek.[22] Een soortgelijke exercitie vinden we terug in de vergelijking van Canadese en Amerikaanse recensies van de romans van Margaret Atwood door de Duitse literatuurwetenschapper Reingard M. Nischik, die daarmee beargumenteert dat 'the variable of nation' – ook in 'tijden' van globalisering – nog steeds een cruciaal is voor de vorming van literaire oordelen.[23]

Waar deze casussen duidelijk inzichtelijk hebben gemaakt *dat* de nationaliteitsvariabele een rol speelt in de evaluatie van culturele producten, kunnen ze ons minder vertellen over *hoe* nationaliteit de wijze van evalueren precies beïnvloedt – daarvoor is het analytisch niveau te particulier, te beperkt tot de specifieke casus. Evalueren Amerikaanse recensenten, om in de casus van Griswold te blijven, altijd vanuit een meer gepolitiseerd, minder autonomistisch kader? Of doen ze dat alleen wanneer romans problematiek thematiseren die diepgeworteld is in de Amerikaanse cultuur? Of gelden de door Griswold opgemerkte verschillen juist alleen voor het werk van Lamming? Om dergelijke vragen te beantwoorden, moeten we het niveau van het particuliere – van individuele romans, van individuele auteurs – juist ontstijgen, en meer algemeen geldende patronen, wetmatigheden en verschillen opsporen. Vanuit een internationaal onderzoeksperspectief analyseert deze scriptie dan ook hoe de nationaliteitsvariabele een rol speelt in de evaluatie van literaire werken in twee (sociaal-)cultureel soortgelijke landen. Daarin vormt een vergelijking tussen de Nederlandse en Vlaamse literatuur een

19 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 17.

20 Zie Hoofdstuk 5.1.3 voor een uitgebreidere toelichting op deze – toch niet onbelangrijke – keuze.

21 Bijvoorbeeld in Amanda Brandellero en Olav Velthuis, "Reviewing art from the periphery. A comparative analysis of reviews of Brazilian art exhibitions in the press", *Poetics* 71 (2018): 58.

22 Wendy Griswold, "The fabrication of meaning: Literary interpretation in the United States, Great Britain, and the West Indies", *American Journal of Sociology* 92, nummer 5 (1987): 1077–1117.

23 Reingard M Nischik, "Margaret Atwood's Reception in Canada and the United States: A Comparative Analysis of North American Reviews of The Blind Assassin", *Contemporary Women's Writing* 11, nummer 3 (2017): 370; Zie ook: Reingard M Nischik, "Reviewing Atwood in Canada and the United States: From (Inter) Nationalism to Transnationalism", in *Comparative North American Studies* (Springer, 2016), 121–178.

uitstekende casus – op een hoger niveau dan de studies van Griswold en Nischik – omdat deze het mogelijk maakt zo veel mogelijk andere factoren, zoals de taal, de besproken romans, de mate van culturele centraliteit, de geografische ligging en welvaart uit te sluiten.

Specifieke aandacht gaat in deze scriptie daarbij uit naar de rol die 'het politieke' speelt in de Nederlandse en Vlaamse literatuurkritiek. Zoals verschillende letterkundigen hebben beargumenteerd, is de geschiedenis van de Vlaamse literatuur nauw verbonden met de Vlaamse beweging.[24] Veel meer dan in Nederland diende literatuur in Vlaanderen een emancipatoir doel: ze stond, aldus Buelens, 'centraal in de Vlaamse ontvoogdingsstrijd'.[25] Tegelijkertijd zien we in Nederland – waar een vergelijkbare sterke politieke inbedding ontbreekt – na de Tweede Wereldoorlog, (mede) onder invloed van het tijdschrift *Merlyn*, juist een voorkeur voor meer werkimmanente, 'autonomistische' literatuuropvattingen. Dit '*Merlyn*-effect' vinden we ook terug in de veranderende aard van de Nederlandse literatuurkritiek na 1945, waar steeds meer aandacht uitgaat naar structurele eigenschappen van literaire werken, en minder naar zogeheten buitenliteraire factoren.[26] Hoewel er niet zelden vanuit gegaan wordt dat de invloed van *Merlyn* ook tot Vlaanderen reikte, zijn de concrete effecten van dit overwegend Nederlandse tijdschrift nog alleen op uitsluitend Nederlandse corpora onderzocht. Dit onderzoek vertrekt vanuit de hypothese dat de nauwe verwevenheid van de Vlaamse Beweging met de Vlaamse literatuur ervoor zorgde dat Vlaamse literatuurcritici literaire werken vanuit een minder autonoom (en meer gepolitiseerd) poëticaal kader beoordelen dan hun Nederlandse collega's – en verwacht dan ook dat het *Merlyn*-effect voor de Vlaamse literatuurkritiek aanzienlijk kleiner is.

Daar speelt ook een tweede sociologische factor mogelijk een rol van betekenis: de latere ontzuiling in Vlaanderen. In Nederland wordt de ontzuiling doorgaans kort na de Tweede Wereldoorlog gesitueerd, met de jaren zestig als het aangewezen epicentrum, waarin de ontzuiling in een stroomversnelling zou raken[27] – al hebben letterkundigen als Anbeek en Brems gesuggereerd dat tekenen van ontzuiling in het literaire veld al vroeger zichtbaar waren.[28] Voor Vlaanderen ligt de zaak aanzienlijk complexer. Daar zien we

24 Brems, "Het centrum als periferie als centrum", 214; Humbeeck, "Een knuppel in het boekenhok: schrijf Vlaamse literatuur, geen Zuid-Nederlandse!"; Buelens, "Gotspe of godsgeschenk? Enkele notities over zin en onzin van een Belgische literatuurgeschiedenis", 89.

25 Buelens, *Van Ostaijen tot heden: zijn invloed op de Vlaamse poëzie*, 89.

26 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 200-201; Marieke Vermeulen, "De dagbladkritiek van vóór en ná Merlyn. Verschillen tussen de literatuurkritische praktijk in 1955 en 1975", *Tijdschrift voor Tijdschriftstudies*, nummer 29 (2011): 18–35.

27 Historicus Hans Righart heeft het in deze context over een (vermeende) 'plotselinge ontzuiling', volgens James Kennedy ging de ontzuiling van christelijke organisaties 'zeer snel', en letterkundigen Frans Ruiter en Wilbert Smulders spreken over 'de grote ontzuiling van de jaren zestig'. Zie: Hans Righart, *De eindeloze jaren zestig. Geschiedenis van een generatieconflict* (De Arbeiderspers, 1995), 31; James C Kennedy, *Nieuw Babylon in aanbouw: Nederland in de jaren zestig* (Boom Koninklijke Uitgevers, 1995), 98; Frans Ruiter en Wilbert Smulders, *Literatuur en moderniteit in Nederland: 1840-1990* (De Arbeiderspers, 1996), 21.

28 Ton Anbeek, "Niet het moment voor experimenten. De twee gezichten van de jaren vijftig", *Een stille revolutie? Cultuur en mentaliteit in de lange jaren vijftig*, 1997, 19–26; Hugo Brems, *Altijd weer vogels die nesten beginnen: geschiedenis van de Nederlandse literatuur, 1945-2005* (B. Bakker, 2006), 36.

rond de jaren zestig weliswaar ook een toenemende ontkerkelijking en modernisering, maar namen de ledenaantallen van zuilorganisaties juist toe.[29] Ook het Cultuurpact van 1973, waarin de aanwezigheid van ideologische en filosofische minderheden in culturele instellingen werd gewaarborgd, kan niet eenduidig in de richting van verdere verzuiling of beginnende ontzuiling geïnterpreteerd worden: enerzijds zorgde het voor 'de verdere pacificatie van de levensbeschouwelijke breuklijn in België',[30] anderzijds werd de verzuiling er juist in 'vastgelegd en officieel geïnstitutionaliseerd'.[31] Historici maken in een Vlaamse context dan ook veelal het onderscheid tussen een 'mentale', of 'geestelijke' ontzuiling, en een veel latere 'institutionele' of 'organisatorische' ontzuiling.[32] Waar de eerste zich vanaf de jaren zestig en zeventig al aantekende, leidde ze pas vanaf de jaren tachtig of begin jaren negentig – afhankelijk van de sector – tot substantiële institutionele verzwakking en hervorming.[33]

Wat betreft de literaire praktijk stelt Hugo Brems dat 1985 een belangrijk kantelpunt in de Vlaamse literatuur markeert.[34] De dominante positie van de drie grote verzuilde literaire tijdschriften in Vlaanderen brokkelde snel af, de ideologische scheidslijnen 'in de literaire praktijk zelf [waren] verregaand vervaagd' – al bleven 'levensbeschouwelijke gevoeligheden' in Vlaanderen anno 1985 nog sterker dan in Nederland.[35] Op grond hiervan kunnen we dus eveneens verwachten dat buitenliteraire factoren – en dan vooral levensbeschouwelijke en morele kwesties – langer hun stempel drukten op de oordelen van Vlaamse critici, en pas later veel later een toegenomen aandacht voor de meer werkimmante aspecten zullen vinden.

Het vervolg van deze scriptie, waarin ik stapsgewijs een antwoord geef op de in deze inleiding geformuleerde vragen, valt uiteen in vier delen en acht hoofdstukken. In het eerste deel schets ik de theoretische en wetenschappelijke achtergrond van deze scriptie. Ik ga eerst uitgebreid in op de fundaties en de achtergrond van het categoriseringsmodel van Linders en Op de Beek, om vervolgens mijn poging deze categorisaties automatisch te reproduceren binnen recente debatten over digitale benaderingen in de geesteswetenschappen te plaatsen. In het tweede deel ga ik specifiek in op de methode en de dataverzameling. Voor het methodische deel is een aanzienlijk deel ingeruimd, en hoofdstuk vier, waarin ik vrij technisch rapporteer hoe ik het categoriseringmodel computationeel heb geïmplementeerd en geëvalueerd,

29 Els de Witte en andere, *Politieke geschiedenis van België: van 1830 tot heden*, E-bookeditie (Manteau, 2016), hoofdstuk V: Van de naoorlogse wederopbouw naar explosieve sixties (1944-1974); Gita Deneckere en andere, *Een geschiedenis van België*, 2012, 198.

30 Deneckere en andere, *Een geschiedenis van België*, 202-203.

31 Witte en andere, *Politieke geschiedenis van België: van 1830 tot heden*, hoofdstuk V.

32 Staf Hellemans, "Verzuiling en ontzuiling van de katholieken in België en Nederland. Een historisch-sociologische vergelijking", *Sociologische Gids* 35, nummer 1 (1988): 50; Witte en andere, *Politieke geschiedenis van België: van 1830 tot heden*; Guido Dierickx, "De verzuiling van belangengroepen: over de kunst van beschrijving en verklaring", in *Tussen bescherming en verovering: sociologen en historici over zuilvorming* (1988), 51.

33 Luc Huyse, "De ontzuiling is nog niet verteerd", *Samenleving en politiek* 20, nummer 2 (2013): 30–39; Witte en andere, *Politieke geschiedenis van België: van 1830 tot heden*, hoofdstuk VI: Samenspel van legitimiteitscrisissen op economisch en institutioneel vlak (1976-2016); Buelens, *Van Ostaijen tot heden: zijn invloed op de Vlaamse poëzie*, 53.

34 Brems, *Altijd weer vogels die nesten beginnen: geschiedenis van de Nederlandse literatuur, 1945-2005*, 486.

35 Ibidem, 484.

kan dan ook tot op zekere hoogte gelezen worden als een afgebakend sub-onderzoek binnen het grotere geheel van deze scriptie. In het derde deel presenteer ik de resultaten van dit onderzoek. In het eerste hoofdstuk daarvan, in hoofdstuk zeven, richt ik mij op de analyse van de aspectcodes van het categoriseringsmodel, en in hoofdstuk acht op de eigenschapscodes. De bevindingen over de gehele linie van deze scriptie synthetiseer ik ten slotte in het vierde en laatste deel: de conclusie.

Deel I

ACHTERGROND

Hoofdstuk 2

LITERATUURKRITIEK VANUIT KWANTITATIEF PERSPECTIEF

Deze scriptie is eerst en vooral – theoretisch, methodologisch – in hoge mate schatplichtig aan de eerdergenoemde proefschriften van Yvette Linders en Esther op de Beek. In dit hoofdstuk zal ik dan ook in de eerste plaats uitgebreid aandacht besteden aan de begripsbepalingen, de representatieve reikwijdte en de theoretische en methodologische veronderstellingen die aan het model van Linders en Op de Beek ten grondslag liggen.

Zowel het proefschrift van Esther op de Beek, getiteld *Een literair fenomeen van de eerste orde*, als Yvette Linders' *Met waardering gelezen* maakt onderdeel uit van het bredere NWO-project *The Best Intentions. Literary Criticism in the Netherlands 1945-2005*. Doel van het project is om de naoorlogse Nederlandse literatuurkritiek *als geheel* – en dus niet aan de hand van enkele case studies – vanuit een kwantitatief perspectief te bestuderen,[36] en door samenwerkingen tussen taalbeheersers (Linders) en letterkundigen (Joosten, Op de Beek) een categoriseringsmodel voor evualatieve uitspraken in literaire kritiek te ontwikkelen.

Een dergelijk categoriseringsmodel is, zoals de onderzoekers zelf ook aangeven, niet nieuw: binnen de letterkundige neerlandistiek hebben onder andere J.J.A. Mooij, H.T. Boonstra en Olf Praamstra al eerder een (theoretische) aanzet tot de categorisering van literairkritische uitspraken gedaan,[37] in Duitsland is er al een langere traditie binnen dit wetenschapsgebied,[38] en ook in de studie van Wendy Griswold wordt er gebruik gemaakt van een handmatig annotatiemodel om literairkritische uitspraken te classifice-

36 Jos Joosten, *The Best Intentions: Literary Criticism in The Netherlands 1945-2005*, https://www.nwo.nl/onderzoek-en-resultaten/onderzoeksprojecten/i/10/3010.html, Geraadpleegd: 2019-05-08.

37 HT Boonstra, "Van waardeoordeel tot literatuuropvatting", *De Gids* 142, nummer 4 (1979): 243–253; JJA Mooij, "De motivering van literaire waardeoordelen", *Tekst en lezer. Opstellen over algemene problemen van de literatuurstudie*, 1979, 253–278; Olf Praamstra, "De analyse van kritieken", *Voortgang. Jaarboek voor de neerlandistiek* 5 (1984): 241–264.

38 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 16; Zie bijvoorbeeld: Lieselotte Schmuck, "Literaturkritik und literarische Wertung. Aspekte einer inhaltsanalytischen Untersuchung deutschsprachiger Romankritik (1945-75)", *Zeitschrift fur Literaturwissenschaft und Linguistik, Beiheft* 12 (1981): 96–115.

ren. Wél nieuw is de grootschalige, longitudinale toepassing van een dergelijk model op een corpus aan Nederlandse romanrecensies, 734 in totaal, verdeeld over vijf kranten en zes pijljaren (1955, 1965, 1975, 1985, 1995 en 2005). Waar, zeker binnen de neerlandistiek, daadwerkelijke toepassingen van zulke modellen uitbleven, is het model van Linders en Op de Beek ontstaan, aangepast en uiteindelijk ook verbeterd in interactie met concrete recensies. Deze grootschalige toepassing, met het volledig gecodeerde recensiecorpus als resultaat, biedt ons uiteindelijk ook de unieke mogelijkheid een dergelijk categoriseringsmodel computationeel te implementeren – zonder deze rijke (en ook: dure) handmatig gecodeerde dataset is dat (vooralsnog) nog niet mogelijk.

2.1 EVUALATIEVE UITSPRAKEN

Het model van Linders en Op de Beek beperkt zich tot de categorisatie van zogeheten 'evualatieve uitspraken' in recensies: uitspraken waarin 'een waardeoordeel te herkennen is'.[39] Daarmee onderscheiden de onderzoekers *evualatieve* uitspraken ('Dit boek is mooi') van *descriptieve* uitspraken, waarin een eigenschap van het boek besproken wordt, maar geen waardeoordeel aan opgehangen wordt ('Dit boek heeft een felrode kaft'). Het onderscheid hiertussen is theoretisch helder, maar in de praktijk veel minder eenduidig. Zo beargumenteert Op de Beek dat een uitspraak zoals 'Dit boek is een historische roman' niet onomstotelijk in een van twee categorieën te plaatsen is:

> Er worden eigenschappen toegekend aan een boek op basis van een interpretatie ervan: een interpretatieve uitspraak dus, al is misschien gewoon de aanduiding die op de kaft van de roman staat door de criticus 'descriptief' overgenomen.[40]

Daaruit concludeert Op de Beek dat het onderscheid tussen een zuiver descriptieve en een evaluatieve uitsprak 'in geen enkel geval empirisch aantoonbaar' is, maar volledig afhankelijk is van de – tekstuele, communicatieve, literair-hiërarchische – context.[41] Daarom worden zowel impliciet als expliciet geformuleerde waardeoordelen in het onderzoek meegenomen, waarbij gold dat wanneer één van de twee onderzoekers een uitspraak als evaluatief interpreteerde, deze ook als zodanig werd gecodeerd.[42]

Een literair waardeoordeel wordt gerechtvaardigd aan de hand van, of gevormd door, bepaalde evaluatiecriteria: de eisen waaraan (goede) literatuur volgens een criticus moet voldoen. Een criticus kan ervoor kiezen deze evaluatiecriteria in zijn recensie (vooraf) te expliciteren, zoals in voorbeeld (1), maar veel vaker – zo blijkt ook uit het pilotonderzoek van Linders en Op

39 Linders, "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005", 19.
40 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 109.
41 Ibidem, 110.
42 Ibidem, 127.

de Beek – formuleert een criticus zijn beoordelingscriteria impliciet, zoals in voorbeeld (2).

(1) Een goede roman zit compositorisch sterk in elkaar.

(2) Dit boek is een compositorische chaos.

Waar in voorbeeld (1) (stellig) geëxpliciteerd wordt dat goede boeken compositorisch goed in elkaar moeten zitten, zou je dat in (2) louter af kunnen leiden uit het feit dat compositie voor deze hypothetische recensent dusdanig belangrijk is dat hij dat als criterium benoemt. Omdat expliciete criteria zoals in voorbeeld (1) – Linders en Op de Beek spreken in deze context over 'metakritische uitspraken' – slechts zelden voorkomen, zeker binnen de beperkte ruimte van een dagbladrecensie, focussen de onderzoekers zich primair op uitspraken van het type (2).[43]

2.2 ASPECTEN EN EIGENSCHAPPEN

In het analyseren van deze evaluatieve uitspraken focussen Linders en Op de Beek zich op het begrip 'evaluatiedomein': 'het onderwerp van evaluatie'.[44] Het gaat Linders en Op de Beek dus vooral om een inhoudelijke beschrijving van literaire evaluaties, waarbij niet zozeer de vorm van evaluatieve uitspraken, of hun retorische legitimiteit, maar 'dat wat in een evaluatieve uitspraak wordt geëvalueerd' het analytische zwaartepunt vormt. Daarbij onderscheiden de onderzoekers twee componenten van een evaluatieve uitspraak: het *aspect* dat geanalyseerd wordt (de stijl, het plot, de dialogen), en de *eigenschap* die daaraan toegekend wordt (helder, humoristisch, emotioneel). De combinatie van aspecten en eigenschappen vormt het evaluatiedomein. Deze bestaat dan ook altijd uit minimaal één aspect en één eigenschap, maar het kan ook zo zijn dat er meerdere eigenschappen worden toegekend aan één aspect, of vice versa.

De aspecten en eigenschappen die toegekend kunnen worden, zijn in het model van Linders en Op de Beek vervolgens opgedeeld in vaste categorieën, '[o]m het analyse-instrument overzichtelijk te houden'.[45] Tabel 2.1 toont de verschillende aspect- en eigenschapscodes die Linders en Op de Beek onderscheiden. Vanwege de beperkte ruimte zal ik in deze scriptie niet uitgebreid op de specifieke definitie van deze categorieën ingaan – de meeste categorieën spreken ook grotendeels voor zich. Voor een verdere specificering met voorbeelden verwijs ik de lezer naar het uitgebreide overzicht dat Linders in haar proefschrift van het categoriseringsmodel geeft.[46]

43 Linders verklaart de lage frequenties van expliciete criteria overigens als volgt: 'Mogelijk verwoorden critici alleen die criteria expliciet waar discussie over bestaat, of juist alleen die criteria die algemeen geaccepteerd zijn'. Zelf vermoed ik eerder een terughoudendheid bij critici met het uitspreken van dergelijke wetmatige, algemeen-geldende evualatiecriteria: voor elk criterium valt wel een wereldklassieker te verzinnen die juist helemaal niet aan dat specifieke criterium voldoet.

44 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 18.

45 Linders, "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005", 52.

46 Zie ibidem, 53-80 en 99-106.

Tabel 2.1: Het uiteindelijke categoriseringsmodel van Linders en Op de Beek.

(a) Aspecten		(b) Eigenschappen	
A	stijl/toon	1	efficiëntie
B	structuur	2	samenhang
C	setting	3	continuïteit
D	plot	4	intensiteit/levendigheid
E	dialogen	5	literaire/intellectuele eigenschappen
F	personages	6	humor
G	thematiek/ideeën	7	helderheid
H	symboliek	8	realiteitsgehalte/geloofwaardigheid
I	uiterlijk en presentatie	9	emotionaliteit
J	auteur/intentie	10	didactische eigenschappen
K	geheel/deel	11	religieuze eigenschappen
L	overig	12	politiek-maatschappelijke eigenschappen
		13	morele eigenschappen
		14	originaliteit
		15	algemeen/overig

2.3 LITERATUUROPVATTINGEN

Met de aandacht voor het coderen van aspecten en eigenschappen van literaire evaluaties wil Op de Beek zich onderscheiden van 'het reconstructieve poëticaonderzoek'.[47] Meermaals benadrukt ze dat haar focus niet ligt op (het reconstrueren van) de poëticale opvattingen van individuele critici.[48] Sterker nog: voor Op de Beek vormen veronderstellingen of literatuuropvattingen van critici überhaupt niet het object van onderzoek. 'Het model', zo stelt ze, 'is bedoeld om evaluaties te ordenen, niet om literatuuropvattingen te reconstrueren'.[49] Het gaat daarbij dus om de manier waarop critici daadwerkelijk in de praktijk evalueren, over de evaluatiecriteria die critici al dan niet bewust gebruiken, en niet over hun achterliggende opvattingen; 'niet wat critici denken, maar wat zij schrijven'.[50] De mate waarin bepaalde aspecten en eigenschappen voorkomen wordt daarom vooral gezien als een indicatie van het feit 'dat die aspecten of eigenschappen een belangrijke plek innemen in het professioneel, publiekelijk en schriftelijk, ofwel gelegitimeerd, evalueren van literaire werken in de dagbladen'.[51]

De voorzichtigheid waarmee Op de Beek aan het analysemodel het vermogen toekent om uitspraken te doen over literatuuropvattingen, is begrijpelijk, en deels terecht. Eerdere modellen om literaire kritiek op structurele wijze te analyseren – in het bijzonder die van Mooij, Boonstra en Praamstra – zijn uitvoerig bediscussieerd en vergeleken, waarbij de discussie vooral draaide om de vraag in hoeverre deze modellen in staat zijn poëticale opvattingen te reconstrueren. Zo is de centrale kritiek van Praamstra op het model van Boonstra dat er geen plaats is voor de analyse van *algemene* uitspraken

47 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 48.
48 Vgl. ibidem, 22, 175 en 249.
49 Ibidem, 22.
50 Ibidem, 18.
51 Ibidem.

in recensies, terwijl deze ons juist bijzonder veel zouden vertellen over literatuuropvattingen en denkbeelden van critici.[52] Harry Bekkering verwijt Praamstra op zijn beurt te veel te focussen op details en nuancering, terwijl het bij literatuuropvattingen vooral zou draaien om 'de hoofdaccenten'.[53] Met een model dat expliciet *niet* dient om individuele poëtica's te reconstrueren, vermijdt Op de Beek gelijkaardige kritiek en discussie.

Dat betekent echter niet dat Op de Beek met haar analysemodel helemaal niets zegt over veranderende literatuuropvattingen. In tegendeel: zoals vermeld in de inleiding, vormt de vraag 'hoe maatschappelijke veranderingen, *poëticale verschuivingen* [mijn cursivering, JV] en literaire ontwikkelingen de wijze van evalueren beïnvloeden' juist één van de voornaamste aanleidingen van het onderzoek.[54] En ook de centrale these van haar onderzoek – dat critici onder invloed van *Merlyn* zich meer toe gingen leggen op structurele analyses van literaire werken – is mijns inziens niet los te zien van veranderende literatuuropvattingen. Daar lijkt Op de Beek zich ook deels bewust van. Zo schrijft ze in haar conclusie: 'Het aantal structuurevaluaties, de categorie die het meest *verbonden* [mijn cursivering, JV] is aan de poëticale opvattingen van Merlyn, bleek in 1985 hoger te zijn dan in andere jaren'.[55] De categorieën in het model dienen dus niet ter reconstructie van poëticale opvattingen, al kunnen deze er wel in meer of mindere mate verbonden aan zijn. Omdat Op de Beek echter bij voorbaat uitsluit dat we met het analysemodel literatuuropvattingen kunnen herleiden, wordt deze complexe relatie tussen het model en literatuuropvattingen – die zich mijns inziens niet alleen in bovenstaand citaat manifesteert, maar door het hele proefschrift impliciet aanwezig is – niet goed uitgewerkt of onderbouwd.

Ik sluit mij bij Op de Beek aan in de zin dat ik in deze scriptie niet geïnteresseerd ben in het reconstrueren van individuele poëtica's. Tegelijkertijd beoog ik met de vergelijking van Nederlandse en Vlaamse literatuurkritiek aan de hand van het analysemodel niet alleen wat te zeggen over de *manier* waarop Nederlandse en Vlaamse critici romans evalueren, maar meen ik dat een vergelijking ons ook iets vertelt over de achterliggende poëticale opvattingen van critici uit Nederland en Vlaanderen.

Eerst een stap terug. Rond de school van Sötemann luidt de gangbare definitie van 'poëtica' als volgt: 'het geheel van opvattingen van een auteur (of groep van auteurs) over literatuur, blijkend uit uitspraken gedaan in en buiten het literaire werk'.[56] Hoewel het begrip door de eeuwen heen een radicale verandering heeft doorgemaakt – van concrete 'leerboeken der dichtkunst' tot een meer abstracte notie van literatuuropvattingen – hebben poëtica's, zo schrijft Wiljan van den Akker in de inleiding van zijn proefschrift, in de twintigste eeuw een onveranderd normatief karakter: 'aan iedere vorm van literair-kritische beschouwing [ligt] een normatief moment

52 Praamstra, "De analyse van kritieken", 243.
53 Harry Bekkering, "Veroverde traditie: de poëticale opvattingen van S. Vestdijk" (proefschrift, [Sl: sn], 1988), 39.
54 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 17.
55 Ibidem, 240.
56 Wiljan Van den Akker, *Een dichter schreit niet: aspecten van M. Nijhoffs versexterne poetica* (Veen, 1985), 10.

ten grondslag'.[57] En in de meeste gevallen zijn deze kritische normen 'niet uitgesproken, maar worden ze zichtbaar in contrast met die van anderen'.[58]

Vanuit deze notie van poëtica wordt de sprong naar het analysemodel van Linders en Op de Beek al enigszins gemakkelijker. Linders en Op de Beek focussen zich primair op de evaluatieve uitspraken waarin de specifieke *normen* waaraan literatuur moet voldoen impliciet blijven, en waar de criticus zich mogelijk niet helemaal van bewust is. Dit maakt inderdaad dat er geen een-op-eenrelatie is tussen de wijze waarop critici evalueren en hun literatuuropvattingen, en we ook niet zozeer kunnen spreken over het reconstrueren van literatuuropvattingen. Maar dit betekent evenmin dat er geen relatie tussen beide is. Waar we met het analysemodel niet dominante literatuuropvattingen kunnen *reconstrueren*, geeft het er wel een sterke *indicatie* van. Wanneer blijkt dat er in Vlaamse recensies consequent meer aandacht is voor de politieke eigenschappen van romans dan in Nederlandse recensies, is dat een duidelijke indicatie van het feit dat Vlaamse critici veel sterker van een politiek-poëticaal kader lezen. Daarmee doe ik geen uitspraken over de specifieke poëticale opvattingen van een individuele criticus, maar probeer ik algemene normen en opvattingen vanuit een cultureel-maatschappelijk kader te verklaren.

57 Van den Akker, *Een dichter schreit niet: aspecten van M. Nijhoffs versexterne poetica*, 9-12.
58 Ibidem.

Hoofdstuk 3

DIGITAL HUMANITIES

Nederlandse en Vlaamse literatuurkritiek zijn – in tegenstelling tot traditioneel literatuurwetenschappelijk onderzoek, waarin veelal individuele werken, genres, auteurs of critici centraal staan – twee kolossale categorieën. Met het automatisch aanleren van het analysemodel van Linders en Op de Beek kiest deze scriptie voor een kwantitatieve en computergestuurde analyse om deze twee categorieën zo accuraat mogelijk te benaderen. Hoewel het analyseren van enkele *case studies* ook een valide methode zou kunnen vormen, biedt de kwantitatieve insteek hier bij uitstek de mogelijkheid om het niveau van het particuliere – van verschillen tussen individuele critici, tussen bepaalde besproken werken – te ontstijgen, en meer algemeen geldende patronen op te sporen en 'hard' te maken. Of, om met een bekend citaat van Franco Moretti te spreken: 'A field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it isn't a sum of individual cases: it's a collective system, that should be grasped as such, as a whole'.[59] Daarmee kan deze scriptie gesitueerd worden binnen het veld van de *Digital Humanities* (DH) – een koepelterm voor computergestuurde benaderingen in de geesteswetenschappen – in het algemeen, en *Computational Literary Studies* (CLS) in het bijzonder.

In dit hoofdstuk positioneer ik mij nader binnen de *Digital Humanities*. Daarbij zal ik mij niet, zoals wel eens gebruikelijk lijkt, toespitsen op het bieden van een uitgebreid overzicht van de geschiedenis en de definitiekwesties van de *Digital Humanities* – van Roberto Busa tot Franco Moretti – maar focus ik mij doelbewust op publicaties die de laatste vijf à tien jaar zijn verschenen. Door razendsnelle technologische ontwikkelingen is het veld immers in beweging, en kunnen computers taal begrijpen, verwerken en contextualiseren op manieren die tien jaar geleden nog niet voor mogelijk werd gehouden – hetgeen meer gedateerde reflecties op de *Digital Humanities*, zeker wanneer deze de limitaties van computationeel onderzoek bediscussiëren, aan relevantie laat inboeten.

59 Franco Moretti, *Graphs, maps, trees: abstract models for a literary history* (Verso, 2005), 1.

3.1 DISTANT READING

Centraal in veel computationele toepassingen binnen de literatuurwetenschap staat de notie van *distant reading*,[60] een term die rond 2000 door Moretti geponeerd is als tegenhanger van het traditionele *close reading*.[61] Moretti's klassieke argument voor *distant reading* – en dit heeft hij op bijzonder veel plekken herhaald – is het idee van *The Great Unread*.[62] Het argument luidt als volgt: Het corpus aan literaire teksten dat een mens in zijn leven kan lezen is beperkt, en behelst, zelfs wanneer je in een bijzonder hoog tempo van een roman per dag leest, altijd maar een fractie van de totale literaire productie – niet meer dan een procent.[63] *Distant reading*, waarbij we boeken niet meer in zijn geheel lezen, maar er telbare eenheden in onderscheiden, biedt daarbij de uitweg: plots kunnen we wel dertigduizend romans in één keer analyseren, en daarmee 'beyond the canon' kijken.[64]

Hoewel de term *distant reading* aan Moretti is toegeschreven, wordt zijn werk en argumentatie mijns inziens onterecht als representatief geacht voor computationele toepassingen in de letterkunde.[65] Moretti schrijft bijzonder polemisch en interpreteert suggestief, formuleert nauwelijks concrete onderzoeksvragen en werkt grotendeels niet met computertechnieken. Ook de notie van *The Great Unread* is mijns inziens allesbehalve waarom computationele technieken in de praktijk worden toegepast. Immers: zeker voor moderne letterkunde geldt dat de onderzoekscorpora nog altijd zeer beperkt worden door auteursrechten. Een banale vergelijking: waar de literatuurgeschiedenis van Hugo Brems ongeveer twaalfhonderd literaire werken citeert,[66] baseren Jautze, Koolen, Van Cranenburgh en De Jong zich voor de vergelijking van *chick lit* en literatuur – geen marginale genres – op een corpus van 32 boeken.[67]

60 Over de representativiteit van Moretti en de praktijk van *The Great Unread* schreef ik eerder in een onderzoeksrapport voor de RMA-cursus *Distant reading* (2018-2019); hoewel een fundamenteel deel is gewijzigd, komen enkele frasen en stellingen (zoals de vergelijking tussen Brems en Jautze et al.) in de eerste twee alinea's van deze paragraaf overeen met dat rapport.

61 Franco Moretti, "Conjectures on world literature", *New left review*, 2000, 54–68.

62 Franco Moretti, *Distant reading* (Verso Books, 2013), 87.

63 Moretti, *Graphs, maps, trees: abstract models for a literary history*, 4.

64 Moretti, "Conjectures on world literature".

65 Zo heeft Ted Underwood ook beargumenteerd dat *distant reading* en de inzet van computationele technieken maar al te vaak als inwisselbaar worden gebruikt, terwijl *distant reading* niet noodzakelijk het gebruik van computationele methoden behelst. De complexiteit van het onderscheid lijkt daarbij te zitten in het feit dat de inzet van computationele middelen wel vaak gepaard gaat met 'reframing the scale of literary inquiry', hetgeen Underwood als kenmerkend voor *distant reading* ziet. Martin Paul Eve's recente *Close Reading with Computers* laat echter eens te meer zien dat digitale methodes niet noodzakelijk ingezet moeten worden om macro-literairhistorische vragen te beantwoorden. In het vervolg van het hoofdstuk hou ik zelf zo veel mogelijk het bovenstaande onderscheid aan. Zie: Ted Underwood, "A Genealogy of Distant Reading.", *DHQ: Digital Humanities Quarterly* 11, nummer 2 (2017); Martin Paul Eve, *Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell's Cloud Atlas* (Stanford University Press, 2019).

66 Dit is een ruwe inschatting – het precieze aantal maakt voor mijn argument niet zozeer uit – op basis van het aantal cursieven in het register. Op de eerste pagina worden 35 literaire werken genoemd. Het register bestaat uit 32 pagina's, dus $35 \cdot 32 = 1120$. Zie: Brems, *Altijd weer vogels die nesten beginnen: geschiedenis van de Nederlandse literatuur, 1945-2005*.

67 Kim Jautze en andere, "From high heels to weed attics: a syntactic investigation of chick lit and literature", in *Proceedings of the Workshop on Computational Linguistics for Literature* (2013), 72–81.

Mijn reden om computationele technieken toe te passen op een groot recensiecorpus sluit eerder aan bij de meer recente ideeën van onderzoekers als Andrew Piper, Matthew Jockers en Ted Underwood.[68] Het gaat mij, zoals ik aan het begin van de paragraaf ook al kort vermeldde, om de mogelijkheid om deze recensies te analyseren op elementen 'that can be generalized, aggregated, and quantified',[69] om op basis van *samples* op herhaalbare, controleerbare en statistisch valide wijze generaliserende uitspraken te doen over de populatie,[70] en *samples* uit verschillende sociale contexten te kunnen contrasteren.[71] Niet de corpusgrootte, maar de mogelijkheid om recensies op herhaalbare wijze te reduceren tot telbare eenheden, is mijn voornaamste argument voor de inzet van computationele technieken. De corpusgrootte en eventuele noties van decanonisering zijn er vervolgens vooral een positief bijeffect van.

Binnen de letterkundige neerlandistiek heeft de tendens richting grootschalige computeranalyses (vooral) geleid tot een reeks artikelen waarin een afstand tussen *close* en *distant reading* wordt gesignaleerd.[72] Ook in een Angelsaksische context wordt een dergelijke afstand gesuggereerd.[73] Het signaleren van zo'n afstand wordt meer dan eens gevolgd door een niet altijd even overtuigende poging het gat tussen *close* en *distant* te dichten, waarin ofwel *close reading* louter wordt gepresenteerd als een manier om enkele opmerkelijke *outliers* in de dataset nauwkeurig te analyseren, ofwel computationele middelen op corpora van bijzonder minimale omvang (vaak: N=1) worden toegepast.[74] Een welkome uitzondering is het artikel van Els Stronks, waarin wordt voorgesteld 'de afstand tussen *distant* en *close* juist zo groot mogelijk [te] maken'.[75] Daar sluit ik mij dan ook volledig bij aan. Zorgen over disciplinaire afzondering van onderzoekers met empirische ambities moeten zonder meer ernstig genomen worden,[76] maar de veronderstelling dat 'de scheiding van vragen en methoden [...] een integratie van digitale methoden in de letterkundige praktijk [bemoeilijkt]', zoals Lucas van der Deijl en

68 Vgl. Andrew Piper, *Enumerations: Data and literary study* (University of Chicago Press, 2018); Matthew L Jockers, *Macroanalysis: Digital methods and literary history* (University of Illinois Press, 2013); Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change* (University of Chicago Press, 2019).

69 Jockers, *Macroanalysis: Digital methods and literary history*, 26.

70 Andrew Piper, "Do we know what we are doing?", *Cultural Analytics*, 2019,

71 Underwood, *Distant Horizons: Digital Evidence and Literary Change*.

72 Vgl. Roel Smeets en Lucas van der Deijl, "Tussen close en distant. Personage-hiërarchieën in Peter Buwalda's Bonita Avenue", *Tijdschrift voor Nederlandse Taal-en Letterkunde* 134 (2018): 123–145; Els Stronks, "De afstand tussen close en distant. Methoden en vraagstellingen in computationeel letterkundig onderzoek", *Tijdschrift voor Nederlandse Taal-en Letterkunde* 129, nummer 4 (2014): 205–214; Karina van Dalen-Oskam, "De stijl van R. (oratie)", 2013,

73 Stephen Ramsay, *Reading Machines: Toward and Algorithmic Criticism* (University of Illinois Press, 2011); Hoyt Long en Richard Jean So, "Literary pattern recognition: Modernism between close reading and machine learning", *Critical Inquiry* 42, nummer 2 (2016): 235–267; N Katherine Hayles, "How we read: Close, hyper, machine", *ADE bulletin* 150, nummer 18 (2010): 62–79.

74 Waarbij het dan de vraag is in hoeverre we hieraan daadwerkelijk het predicaat *distant reading* kunnen toekennen (zie ook mijn eerdere voetnoot over het onderscheid tussen distant reading en de inzet van computationele methoden).

75 Stronks, "De afstand tussen close en distant. Methoden en vraagstellingen in computationeel letterkundig onderzoek".

76 Vgl. Karina van Dalen-Oskam, "Ballingschap der empirici", *Tijdschrift voor Nederlandse Taal-en Letterkunde* 125, nummer 2 (2011): 122–125.

Roel Smeets in een recent artikel schrijven, lijkt mij niet noodzakelijk waar.[77] Juist door radicaal andere methoden te gebruiken en daarmee andere vragen te stellen, juist door te differentiëren in plaats van te assimileren, kan *distant reading* haar toegevoegde waarde in de letterkundige praktijk laten blijken. Integratie vindt dan niet plaats wanneer onderzoekers blijven navelstaren op de vermeende afstand tussen *close* en *distant*, wanneer ze vragen stellen die niet verder reiken dan 'Kan de ook computer X?', maar wanneer ze – vrij evident – inspelen op actuele debatten en vraagstukken binnen de neerlandistiek – zij het met een unieke invalshoek.[78]

3.2 RECENTE KRITIEK

Met de groeiende populariteit van de inzet van digitale technologieën is het niet verrassend dat er ook de nodige kritiek op de *Digital Humanities* is geuit. Deze kritiek richt zich aanvankelijk vooral op de ideologische motieven en redenaties van het veld. Zo bekritiseert Jesse Rosenthal de natuurwetenschappelijke pretenties van het veld, beschouwen Allington, Brouillette en Golumbia de *Digital Humanities* als een 'facilitation of the neoliberal takeover of the university', en richt Alan Lui zijn pijlen op het gebrek aan culturele kritiek in computationele benaderingen.[79] Meer recentelijk is echter, naar aanleiding van het bijzonder felle artikel 'The Computational Case against Computational Literary Studies' van Nan Z. Da, een meer methodologisch georiënteerd debat ontstaan over de statistische robuustheid van computationele benaderingen van literatuur.[80]

Da meent dat papers in CLS op te delen zijn in twee categorieën: 1) papers die niet-significante bevindingen als bevindingen presenteren, en 2) papers die verkeerde conclusies trekken uit significante bevindingen. De hoofdoorzaak van deze problematiek is volgens haar dat CLS-onderzoekers zich alleen bezighouden met het analyseren van woordfrequenties, met het tellen hoe vaak bepaalde woorden voorkomen:

> [A]ll the things that appear in CLS—network analysis, digital mapping, linear and nonlinear regressions, topic modeling, topology, entropy—are just fancier ways of talking about word frequency changes. Breaking down CLS mistakes will clarify why, despite the fact that different semantic and syntactical tagging methods have existed since the 1970s, CLS tends to stick to coun-

77 Smeets en Deijl, "Tussen close en distant. Personage-hiërarchieën in Peter Buwalda's Bonita Avenue", 128.

78 Een uitstekend voorbeeld hiervan vind ik eigenlijk juist een artikel van grotendeels dezelfde auteurs waarin de diversiteit onder Nederlandse romanpersonages onder de loep wordt genomen. Zie: Lucas van der Deijl en andere, "Mapping the Demographic Landscape of Characters in Recent Dutch Prose: A Quantitative Approach to Literary Representation", *Journal of Dutch Literature* 7, nummer 1 (2016).

79 Jesse Rosenthal, "Introduction:"narrative against data"", *Genre* 50, nummer 1 (2017): 1–18; Daniel Allington en andere, "Neoliberal tools (and archives): A political history of digital humanities", *LA Review of Books*, 2016, Alan Liu, "Where is cultural criticism in the digital humanities?", 2012, 490–509.

80 Nan Z Da, "The Computational Case against Computational Literary Studies", *Critical Inquiry* 45, nummer 3 (2019): 601–639.

> ting words and is, in an even more limited sense, forced to find many of its significances by tweaking stop words.[81]

En daardoor zijn computationele benaderingen inherent reductionistisch: verschillen en overeenkomsten tussen complexe, hoog-dimensionale categorieën worden louter verschillen in woordfrequenties. En dat terwijl verschillen in woordfrequenties vrijwel onvermijdelijk zijn: 'data-mine something and you will always find significant associations'.[82] Ofwel: in corpora van miljoenen woorden zijn er altijd wel enkele woorden die in de ene tekstcollege niet, en in de andere tekstcollectie wel voorkomen. Met het negeren van syntax, semantiek, woordpositie en context, met het reduceren van hoogdimensionele teksten naar visualiseerbare datapunten, verliezen we, zo meent Da, een hoop informatie, hoewel ze geen voorbeelden geeft van het gevolg hiervan. Desalniettemin luidt haar conclusie stellig: 'CLS has no ability to capture literature's complexity'.[83]

De bijdrage van Da is met een gelijkaardige felheid bekritiseerd. Zo wijst Andrew Piper terecht op de hypocrisie van haar argument: Da pleit voor robuuste statistische analyses, maar maakt zelf gebruik van een selectieve lezing van computationeel-literatuurwetenschappelijke artikelen, om daar vervolgens 'the broadest, most sweeping claims imaginable' mee te maken, een werkwijze die Piper cynisch plaatst in 'the time-honoured traditions of selective reading from the field of literary criticism'.[84] Een selectieve lezing die bovendien, zoals Fotis Jannidas tegenwerpt, alleen bestaat uit de bespreking van Noord-Amerikaanse onderzoekers.[85] Ook Da's kritiek op CLS' focus op woordfrequenties wordt verworpen. Niet alleen is de stap van netwerkanalyses en digitale cartografie naar woordfrequenties zoals ze die in het bovenstaande citaat suggereert erg groot, ook zijn er voldoende voorbeelden van onderzoekers die zich *juist* met syntactische structuren of sequentiële taalmodellen bezighouden, maar die niet in Da's selectie naar voren komen.[86]

3.3 PRAGMATISME EN PLAATSBEPALING

Hoewel de tekst van Da aan constructiviteit kan winnen, zijn niet alle punten die ze aanstipt onterecht. Ik verkeer (logischerwijs, wellicht) niet in de veronderstelling dat CLS een 'invalid field' is.[87] Evenmin veronderstel ik dat alle computationele analysetechnieken uiteindelijk niets anders zijn dan verschillende manieren waarmee we woordfrequenties kunnen meten. Tegelijkertijd kunnen we niet ontkennen dat woordfrequenties nu eenmaal nog altijd een groot onderdeel zijn van de computationeel-letterkundige praktijk.

81 Da, "The Computational Case against Computational Literary Studies", 607.
82 Ibidem, 609.
83 Ibidem, 634.
84 Piper, "Do we know what we are doing?", 2.
85 Fotis Jannidis, "On the perceived complexity of literature. A response to Nan Z. Da", *Cultural Analytics*, 2019,
86 Michael Gavin, "Is there a text in my data? (Part 1): On Counting Words", *Cultural Analytics*, 2019, Jannidis, "On the perceived complexity of literature. A response to Nan Z. Da".
87 Piper, "Do we know what we are doing?", 2.

Dat betekent niet noodzakelijk dat woordfrequenties niet informatief zijn. In tegendeel: de populariteit van woordfrequenties binnen CLS kan ook deels verklaard worden vanwege het feit dat ze zulke belangwekkende resultaten opleveren. Maar zuiver pragmatisme speelt ook een rol. Woordfrequenties zijn in hoge mate taalonafhankelijk, en de *tools* om woordfrequenties en methodologische varianten te analyseren – *Mallet*, R's *Stylo*, *AntConc* – zijn daarom ook snel, goed ontwikkeld, gebruiksvriendelijk, en vragen om minimale aanpassingen van de datasets.

Veel van deze woordfrequentiemethodes vinden hun oorsprong in de informatica, kunstmatige intelligentie en de computationele taalkunde, en zijn (mede) daarom nog weinig toegespitst op de wensen van geesteswetenschappers. Zo merkt Da terecht dat er geen echte manier is om de accuraatheid van *topic modeling* (hier specifiek: het LDA-algoritme) te meten:

> There is no real way to measure the accuracy of the topics found since LDA's recall depends on having true classes of topics arrived at through human decision making. Its utility is most observable in circumstances where recall and precision really do not matter very much, as with content-based recommendation systems (such as Facebook advertising products to its users).[88]

Waar een dergelijke beperking nauwelijks een rol speelt in een praktische toepassing van LDA, is de mogelijkheid om de validiteit van een model te meten in een wetenschappelijk context bij uitstek relevant. Dat is ook één van de redenen waarom ik geen *topic modeling*, zoals deze techniek conventioneel wordt toegepast binnen de Digital Humanities,[89] of gelijkaardige woordfrequentieanalyses gebruik om mijn hoofdvraag te beantwoorden. Wanneer we zouden vinden dat een *topic* rondom politieke woorden – of simpelweg het woord 'politiek' – na 2000 steeds dominanter in ons corpus wordt, is het moeilijk om kwantitatief te valideren of deze hypothetische stijging verklaarbaar is doordat recensenten romans meer vanuit een politiek-poëticaal lezen (en dus: de politieke eigenschappen van romans *evalueren*), of dat de romans zelf meer politieke onderwerpen aan de orde brengen die de critici zuiver *descriptief* bespreken.[90] Met andere woorden: we weten nooit zeker of we meten wat we willen weten. In veel gevallen tonen andere onderzoekers zich duidelijk bewust van dergelijke beperkingen, maar toch lijkt pragmatisme over het algemeen de doorslag te geven.

88 Da, "The Computational Case against Computational Literary Studies", 625.

89 Bijv. Christof Schöch, "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama.", *DHQ: Digital Humanities Quarterly* 11, nummer 2 (2017); Matthew L Jockers en David Mimno, "Significant themes in 19th-century literature", *Poetics* 41, nummer 6 (2013): 750–769.

90 Hierbij zet ik mij niet zozeer af tegenover *topic modeling* als *techniek* om *features* te generen voor classificatie, of varianten op LDA zoals Labeled-LDA, maar eerder tegenover een veelvoorkomende analyse*methode* in de *Digital Humanities* waarbij er in de toepassing van LDA nadrukkelijk niet gebruik wordt gemaakt van een gelabelde dataset (zie voorbeelden in de voetnoot hierboven). In die zin zet ik mij ook af tegen een *unsupervised* Nearest Neighbors-clusteranalyse op de evenzeer *unsupervised* gegenereerde BERT-embeddings van zinnen in literaire recensies, waar min of meer dezelfde problemen voor gelden, maar het gebruik van een dergelijke methode is – naar mijn weten – aanzienlijk minder conventioneel binnen de DH.

De gestage volwassenwording van *Computational Literary Studies* als discipline houdt mijns inziens echter ook in dat we deels van dat pragmatisme af moeten stappen. Deze scriptie kan dan ook gelezen worden als poging om een computationele methode te ontwikkelen die relevante categorisaties afkomstig uit letterkundig onderzoek op een grote schaal kan reproduceren, om een valideerbaar computationeel model te ontwikkelen dat volledig is toegespitst op de analyse van *Nederlandstalige* literaire recensies. En zoals ik in hoofdstuk 4 uitgebreid zal bespreken, baseer ik dit model niet louter op basis van significante woordfrequenties, maar vergelijk ik verschillende manieren van tekstrepresentaties waarin ook semantische en syntactische informatie verwerkt zit.

Deel II

METHODE

Hoofdstuk 4

DE CLASSIFICATIE VAN RECENSIES

In dit hoofdstuk staat de ontwikkeling en evaluatie van het methodologische fundament van deze scriptie – een automatisch classificatiesysteem voor literaire recensies gebaseerd op het categoriseringsmodel van Linders en Op de Beek – centraal. Ik beschrijf de stappen die ik heb ondernomen om dit model aan de computer te leren, de keuzes die ik daarbij heb gemaakt, en de manier waarop die keuzes de resultaten van deze scriptie beïnvloeden.

De technieken die ik in dit hoofdstuk gebruik en bespreek zijn vooral afkomstig uit het veld van *Natural Language Processing* (NLP) – een wetenschapstak op het raakvlak van computerwetenschappen, kunstmatige intelligentie en taalkunde. Onderzoekers in NLP houden zich bezig met automatische taalverwerking in de breedste zin van het woord, van vertaalmachines, spraakherkenners tot het automatisch ontleden van zinnen. Ook het classificeren van recensies vormt voor veel NLP-onderzoekers één van de kerntaken.[91] Dergelijke classificaties beperken zich echter vrijwel uitsluitend tot de analyse van sentiment in recensies, tot het voorspellen van het aantal sterren – of bolletjes – aan de hand van de tekst van recensie.[92] De allerminste geringe veronderstellingen die aan dergelijke modellen ten grondslag liggen daargelaten, concentreert deze scriptie zich veel minder op de uiteindelijke evaluatie (positief/negatief), en veel meer op welke aspecten en eigenschappen in recensies genoemd worden. Dit vereist dan ook een ander, meer specifiek – en complexer – classificatiemodel. Ik gebruik en vergelijk daarbij de werking van zowel zogeheten *state-of-the-art* taalmodellen (o.a. BERT) als meer gevestigde classificatietechnieken (zoals logistische regressie).

91 Vlg. Jacob Devlin en andere, "Bert: Pre-training of deep bidirectional transformers for language understanding", *arXiv preprint arXiv:1810.04805*, 2018, Zhilin Yang en andere, "XLNet: Generalized Autoregressive Pretraining for Language Understanding", *arXiv preprint arXiv:1906.08237*, 2019, Benjamin van der Burgh en Suzan Verberne, "The merits of Universal Language Model Fine-tuning for Small Datasets–a case with Dutch book reviews", *arXiv preprint arXiv:1910.00896*, 2019,

92 Zo presenteren en testen Van de Burgh en Verberne een gelabelde dataset van 110k boekrecensies van boekensite *Hebban*, waarbij de labels die moeten worden voorspeld het aantal sterren dat de auteur van de recensie aan het boek heeft gegeven reflecteren. Burgh en Verberne, "The merits of Universal Language Model Fine-tuning for Small Datasets–a case with Dutch book reviews".

4.1 DATAVERZAMELING

Voor het ontwikkelen van het automatisch classificatiesysteem voor literaire recensies maak ik gebruik van het compleet gecodeerde recensiecorpus van Linders en Op de Beek. De meest recente versie van het recensiecorpus is verkregen via de onderzoekers en is vooralsnog niet publiekelijk – of voor onderzoekers – digitaal raadpleegbaar. Het aangeleverde bestand is opgemaakt in MAXqda-plus,[93] een softwareprogramma voor kwalitatieve dataanalyse.[94] Hoewel Linders in een softwarereflectieartikel overwegend positief is over de onderzoekstechnische mogelijkheden van het programma, bleek de software en dito bestandsformaat allerminst geschikt voor onderzoekers met meer computationeel-georiënteerde ambities.[95] Zo is het – volgens MAXqda overigens nu al verouderde – bestand alleen werkbaar te openen met een betaalde licentie voor het programma, geïnstalleerd op een Windows-computer (andere besturingssystemen werkten niet of zijn überhaupt niet beschikbaar), en kan er bij het exporteren alleen voor gekozen worden om ofwel alle gecodeerde *passages* (in XSLT-formaat), ofwel alle *teksten* – in RTF-formaat, zonder coderingen – te exporteren, maar is er geen mogelijkheid om de volledige gecodeerde teksten eenvoudig te exporteren. Dat laatste is vanuit commerciële overwegingen logisch – MAXqda maakt het gebruikers zo moeilijk mogelijk om over te stappen naar een (open-source) concurrent. Maar juist dergelijke commerciële overwegingen illustreren wat wetenschapsoftware niet moet zijn: een achter licenties gesloten ecosysteem.

Om het onderzoek in deze scriptie van een meer open en reproduceerbaar karakter te voorzien, zijn alle teksten in het recensiecorpus omgezet naar FoLiA XML, een open dataformaat dat is ontwikkeld door onderzoekers van de Radboud Universiteit Nijmegen, en specifiek bedoeld is voor taalkundige annotatie.[96] FoLiA XML-bestanden zijn zonder externe of gelicentieerde software te openen en raad te plegen, en kunnen programmatisch eenvoudig benaderd worden via XML-parsers. Voor het omzetten van de annotaties heb ik zowel de gecodeerde passages als de volledige ongecodeerde teksten uit MAXqda geëxporteerd. Vervolgens is voor elke gecodeerde passage op geautomatiseerde wijze gezocht naar de positie van die specifieke passage in de volledige recensietekst.[97] De teksten met gekoppelde annotaties zijn via het Python-pakket `PyNLPl`[98] omgezet naar FoLiA-formaat.

93 VERBI Software, *MAXQDA 2018* (Berlijn, Duitsland: VERBI Software, 2017), https://www.maxqda.com.

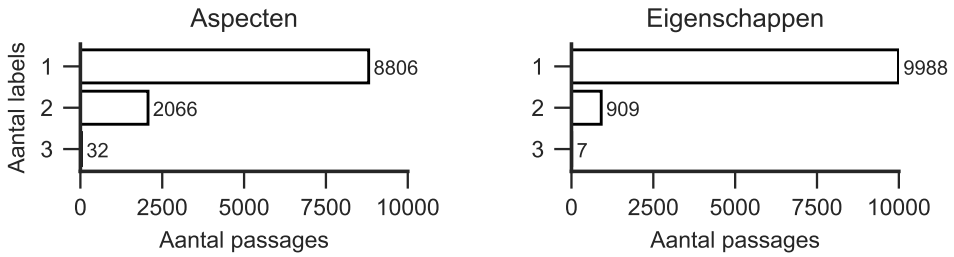
94 'Kwalitatief' moet hier in een meer sociaalwetenschappelijke zin worden opgevat (het handmatig lezen en annoteren van teksten, om deze vervolgens de kwantificeren), niet als de traditioneel-geesteswetenschappelijke analysemethode

95 Yvette Linders, "TS Tools", *Tijdschrift voor Tijdschriftstudies*, nummer 32 (2012): 191–196.

96 Maarten van Gompel en Martin Reynaert, "FoLiA: A practical XML format for linguistic annotation–a descriptive and comparative study", *Computational Linguistics in the Netherlands Journal* 3 (2013): 63–81.

97 Omdat leestekens en alinea-eindes op een andere wijze gecodeerd zijn in het XSLT-bestand met de gecodeerde passages dan in de RTF-teksten, was het vinden van de plek van de passages in de tekst minder eenvoudig dan gedacht. Als een passage niet gevonden kon worden in de tekst, werden eerste alle leestekens uit die passage verwijderd. Wanneer dat nog steeds geen *match* opleverde, werd het eerste en het laatste woord uit de passage verwijderd.

98 https://github.com/proycon/pynlpl



Figuur 4.1: Aantal aspect-gecodeerde en eigenschap-gecodeerde labels per passage

Het gecodeerde recensiecorpus van Linders en Op de Beek bestaat uit 734 gecodeerde dagbladrecensies, met 10.904 verschillende gecodeerde passages waaraan in totaal 24.861 codes zijn toegekend, bestaande uit 13.034 aspectcodes en 11.827 eigenschapscodes. Deze coderingen zijn toegekend aan niet-gefixeerde sequenties van woorden, variërend van enkele woorden, volledige zinnen tot meerdere paragrafen. In het huidige formaat kunnen de gecodeerde passages daarbij overlappen. Zo is aan de volgende passage de code *geheel/deel* toegekend:

> Schaduwkind van P.F. Thomése is zo'n boek. Het is (bijna) onleesbaar, zo afschuwelijk aangrijpend. Dat komt niet alleen door de scènes, maar ook, of misschien vooral, door de taal van Thomése. Hélena, ontbeert de taal van wanhoop.

Terwijl de code *stijl/toon* apart is toegekend aan alleen het laatste deel van deze passage, aan 'door de taal van Thomése. Hélena, ontbeert de taal van wanhoop'. De tekstuele reikwijdte van de gecodeerde passage is in dit formaat dus exact afgestemd op de toegekende codering. Dit maakt dat relatief weinig passages meerdere labels hebben (zie Figuur 4.1): slechts 32 van de 13.034 als aspect gecodeerde passages (minder dan 0.3%) hebben meer dan twee labels.[99]

Hoewel deze niet-gefixeerde manier van annoteren evidente voordelen biedt (onderzoekers kunnen bijvoorbeeld in één opslag zien naar welk specifiek zinsdeel, welke specifieke passage de codering terugverwijst), is een dergelijke werkwijze (vooralsnog) vanuit een computationeel perspectief te complex. Recente zogeheten *state of the art*-modellen voor tekstclassificatie uit de *Natural Language Processing* zijn ruwweg geconcentreerd rond drie smaken van tekstclassificatie: 1) de classificatie van woorden, 2) van zinnen, en 3) van documenten, waarbij vrijwel elke tekstvorm die niet onder smaak 1) of 2) valt als 'document' wordt gezien. Weliswaar zijn er enkele populaire classificatietaken waarin een van deze drie niveaus strikt genomen overschreden wordt – zoals bij het herkennen van eigennamen, waarin de naam *A.F.Th. van der Heijden* als één entiteit wordt gezien – maar over het algemeen veronderstellen tekstclassificatiemodellen een vaste tekstuele reik-

99 Op de Beek geeft dan ook aan dat er bij het coderen 'de voorkeur werd gegeven aan het maken van een keuze voor één aspect en één eigenschap', al 'hebben sommige evaluerende uitspraken onmiskenbaar betrekking op meerdere categorieën uit het model'. Zie: Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 128.

wijdte. In dit onderzoek beschouw ik het herkennen en categoriseren van evaluatieve uitspraken in recensies dan ook als een zinsclassificatieprobleem.

Het omvormen van het niet-gefixeerde annotatiemodel van Linders en Op de Beek naar een zinsclassificatieprobleem kan enerzijds als een methodische beperking van deze scriptie worden gezien: mijn voorgestelde model zal – hoe precies ook – niet in staat zijn de handmatige annotaties van Linders en Op de Beek op exact dezelfde wijze te reproduceren. Ook gaat het gepaard met een verlies aan informatie: wanneer er in een (lange) zin bijvoorbeeld meerdere aparte structuurevaluaties staan, tellen we deze in een zinsclassificatieconfiguratie maar eenmaal. Tegelijkertijd valt er ook wel degelijk iets te beargumenteren voor annotaties op zinsniveau. Een uitspraak in de trant van 'een niet-gefixeerde sequentie van woorden X in document d_i die begint bij word w_i en eindigt bij w_j wordt geassocieerd met een set labels Y uit de labelset L' is vanwege het precies aanwijzen van de *woordelijke* sequentie dan wel meer specifiek dan '*zin* s_i in document d_i wordt geassocieerd met een set labels $Y \subseteq L'$, maar in essentie ook meer arbitrair. Immers: op basis waarvan bepaal je w_i en w_j? In natuurlijke taal: hoe bepaal je waar de gecodeerde sequentie van woorden begint en eindigt? Zo zagen we in de hiervoor geciteerde passage dat de code *stijl/toon* is toegekend aan het deel vanaf het woord 'door'. Maar je zou mijns inziens evengoed kunnen beargumenteren dat de passage over de stijl van Thomése al begint bij het voegwoord, bij 'maar', of zelfs eerder, bij 'Dat komt niet alleen door'. Door uit te gaan van de structuur van de tekst, van de zinsordening die de recensent heeft gemaakt, biedt het classificeren op zinsniveau hier een uitkomst. Want waar je er over kan twisten waar de code *stijl/toon* in deze passages precies begint, is het aanzienlijk lastiger om te beargumenteren dat deze zin helemaal niet de stijl van Thomesé bespreekt.[100]

Om de geannoteerde passages van Linders en Op de Beek om te zetten naar annotaties op zinsniveau, heb ik allereerst de zinnen (en de woorden in die zinnen) gesplitst met *Ucto*, een *tokenizer* die afgesteld is op de Nederlandse taal.[101] Vervolgens ben ik via `PyNLPl` zin voor zin, woord voor woord door de recensieteksten heengegaan. Wanneer minimaal één woord in deze zin voorkwam in een gecodeerde passage, kreeg de zin ook deze codering. De tabellen hieronder illustreren deze transformatie. De bovenstaande tabel representeert de oorspronkelijke dataset D; de onderste tabel de uiteindelijke, getransformeerde dataset D'.

Deze getransformeerde dataset bestaat in totaal uit 27.990 zinnen, waarvan er 10.941 zinnen één of meer labels hebben. Het totaal aantal codes is, doordat sommige gecodeerde passages uit meerdere zinnen bestaan die nu zijn opgesplitst per zin, enigszins toegenomen in vergelijking met de oorspronkelijke dataset: de getransformeerde dataset telt in totaal 30.938 codes, bestaande uit 15.764 aspectcodes en 15.174 eigenschapcodes. Omdat labels van passages die tot dezelfde zin behoren zijn samengevoegd, is ook het aantal labels per passage toegenomen (zie Figuur 4.2): waar een passage eerst

100 Bijkomend voordeel is dat zinnen aanzienlijk makkelijkere kwantificeerbare en vergelijkbare eenheden zijn dan niet-gefixeerde sequenties van woorden. Zowel Linders en Op de Beek gebruiken absolute frequenties van aspect- en evaluatiedomeinen in hun analyse, maar corrigeren daarbij alleen bij het aantal evaluatieve uitspraken voor de lengte van deze passages.

101 Zie `https://languagemachines.github.io/ucto/`

Tabel 4.1: Voorbeeldtransformatie

D		
$w_{1:6}$	Tomas Ross vertelt zijn spannende oorlogsverhaal	$\{D\}$
$w_{7:9}$	in hoog tempo	$\{A\}$
$w_{10:14}$	met treffende historische details en waarheden	$\{C\}$

$\downarrow$

D′		
s_1	Tomas Ross vertelt zijn spannend oorlogsverhaal in hoog tempo met treffende details en waarheden	$\{D, A, C\}$

met maximaal drie labels geassocieerd was, zijn er nu zinnen met meer dan zes labels.

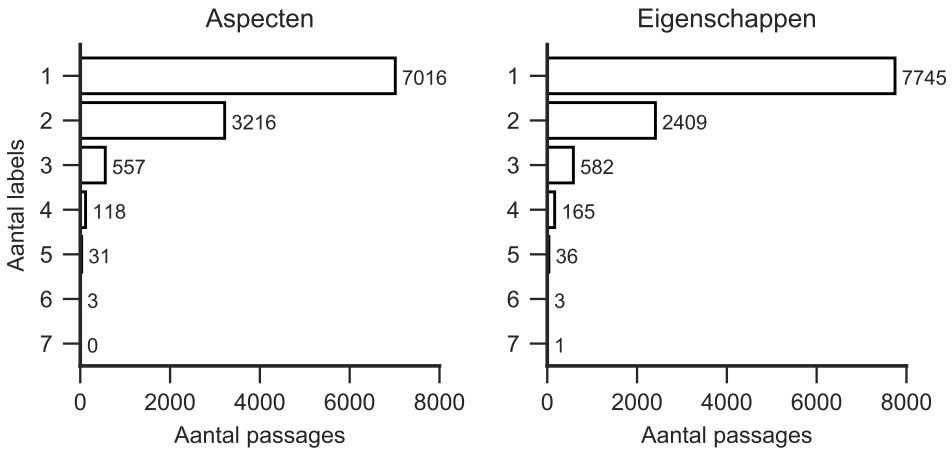
De hierboven voorgestelde datatransformatie bleek in een experimentele fase van het onderzoek echter niet voor elke annotatie in het recensiecorpus even vanzelfsprekend als het voorbeeld in Tabel 4.1 suggereert. Zo werd de als *geheel/deel* gecodeerde passage 'Verwarrend? Nee, allerminst. *Karaktermoord* is verhelderend informatief – bij vlagen zelfs te informatief' opgesplitst in drie aparte zinnen, waarbij niet alleen het laatste deel, maar ook 'Verwarrend?' en 'Nee, allerminst' de code *geheel/deel* apart toegekend kregen. In de context van zinsclassificatie vormen dergelijke elliptische zinnen louter ruis – als je alleen de zin 'Nee, allerminst' neemt, kan je daaruit immers niet herleiden dat ze over de algehele aard van *Karaktermoord* gaat. De keuze om recensies op zinsniveau te classificeren – en dus expliciet niet op woord- of tekstniveau – betekent ook dat de coderingen logisch te herleiden moeten zijn aan de hand van de afzonderlijke zin. Aangezien dit evident niet elke zin het geval bleek, is in dit onderzoek ervoor gekozen de dataset voor dergelijke gevallen te corrigeren. Daarvoor heb ik, met behulp van de *open source* tekstannotatie-tool *Doccano*,[102] alle reeds gecodeerde zinnen handmatig gecontroleerd.

Elke gecodeerde zin werd door het programma in compleet willekeurige volgorde aan mij gepresenteerd, waarbij ik vervolgens heb bepaald of de coderingen voor deze afzonderlijke zin nog steeds stand houdt. Zodoende heb ik 2201 zinnen geïdentificeerd waarvan de labels na het splitsen niet meer kloppen. De dataset telt daarmee uiteindelijk 8713 zinnen met een of meerdere labels.

4.2 PROBLEEMBESCHRIJVING

Waar ik in de voorgaande paragraaf al kort heb toegelicht waarom ik de recensieclassificatietaak beschouw als een zinsclassificatieprobleem, richt ik mij nu op een formelere en meer gedetailleerde beschrijving van het classificatieprobleem. Met andere woorden: wat betekent het nu – in een computa-

102 Beschikbaar via https://github.com/chakki-works/doccano



Figuur 4.2: Aantal labels per passage in de getransformeerde dataset

tionele context – precies om het model van Linders en Op de Beek aan de computer te leren?

Allereerst is het preciezer om niet te spreken over één classificatieprobleem, maar over *meerdere* classificatieproblemen. Het model van Linders en Op de Beek behelst voor menselijke codeurs in essentie drie beslissingsmomenten. Ten eerste moet een codeur besluiten een zin al dan niet als een evaluatieve uitspraak te coderen. Ten tweede moeten er een of meerdere apsectcodes toegekend aan een evaluatieve uitspraak, waarbij gekozen kan worden uit 12 labels, met in totaal $2^{12} - 1 = 4095$ verschillende mogelijke combinaties. Ten derde moet besloten worden welke eigenschapscodes aan de evaluatieve passage worden toegekend, waarbij ditmaal gekozen wordt uit 15 labels, met in totaal maar liefst $2^{15} - 1 = 32767$ mogelijke combinaties. Ook Linders en Op de Beek beschouwen deze drie beslissingsmomenten in hoge mate onafhankelijk van elkaar: bij de bespreking van de interbeoordelaarsbetrouwbaarheid wordt de intersubjectiviteit van deze stappen individueel getoetst.[103]

Hoewel het theoretisch mogelijk is de computer de drie beslissingsmomenten in één keer uit te laten voeren,[104] kies ik er in deze scriptie voor zo dicht mogelijk bij het menselijke beslissingsproces te blijven, en de recensieclassifietaak als een drieledig classificatieprobleem te beschouwen. Dit houdt in dat ik voor elke stap in het beslissingsproces een aparte *classifier* train, die – overeenkomstig met de onderlinge evaluatie van Linders en Op de Beek – ook apart geoptimaliseerd, geëvalueerd en getoetst zal worden. Ook deze keuze is uiteraard niet geheel zonder problemen: het gevaar van een dergelijke benadering is dat zinnen die onterecht als evaluatief geclassificeerd worden, ook aspect- en eigenschapscodes toegewezen krijgen, en fouten van het eerste model op die manier doorsluizen naar het tweede en het derde model. Wanneer je voor één classificatiemodel zou kiezen dat de aan- of afwezigheid van de zevenentwintig codes tegelijkertijd voorspelt, en

103 Linders, "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005", 114-123.

104 Vgl. Folgert Karsdorp, *Retelling Stories. A Computational-Evolutionary Perspective* ([Sl: sn], 2016), 37-38.

een zin niet noodzakelijk een label hoeft de krijgen, zou je *dergelijke* fouten voorkomen. Maar omdat de dataset die er, in ons geval, voor nodig zou zijn een dergelijk model te trainen dusdanig complex, veelzijdig, ongebalanceerd en *sparse* zou zijn, en er daardoor tal van andere classificatieproblemen optreden, is het drieledige classificatiemodel mijns inziens de meest veilige keuze.

Het *herkennen* van evaluatieve uitspraken in recensies beschouw ik, allereerst, als een binair zinsclassificatieprobleem, waar het doel is aan elke zin s_i in document (recensie) d_i een binair label $y_i \in \{0, 1\}$ toe te kennen, waarbij geldt dat s_i een evaluatieve uitspraak bevat wanneer y_i de waarde van 1 aanneemt, en geen evaluatieve uitspraak bevat wanneer y_i gelijk is aan 0. Dit classificatieprobleem vraagt een minimale aanpassing van de huidige dataset, waarbij elke zin die minimaal één label bevat de waarde 1 krijgt, en zinnen zonder label de waarde 0.

Het *toekennen* van zowel aspecten als eigenschappen aan evaluatieve uitspraken benader en evalueer ik in deze scriptie in eerste instantie als een multi-label rankingsprobleem. Het doel is hierbij om voor elke zin s_i in document d_i een ranking te voorspellen van alle labels uit een finiete set labels $L = \{\lambda_1, ..., \lambda_c\}$, waarbij geldt dat labels die hoger in de ranking staan relevanter zijn voor s_i dan labels die lager in de ranking staan. Waar in reguliere classificatie een zin s_i eenduidig geassocieerd wordt met een set labels Y, werkt het voorspellen van een ranking veel minder deterministisch in de zin dat de computer per definitie geen labels *uitsluit*, maar alleen op volgorde van relevantie zet.[105] Of, zoals Folgert Karsdorp dat formuleert in de context van het ranken van motieven in volksverhalen: 'by taking a ranking approach I do not *impose* a classification on a story, but rather, I provide a "motif recommendation list"from which scholars can select the motifs that apply'.[106]

Ik kies voor een multi-label rankingbenadering, en niet voor multi-label classificatie, om twee redenen. Ten eerste hebben zowel latere studenten/onderzoekers die met het model aan de slag zijn gegaan,[107] als de recensenten van beide proefschriften,[108] de fluïditeit van meerdere aspect- en eigenschapscategorieën benadrukt. Ook de weliswaar redelijke, maar 'nog niet optimale' interbeoordelaarsbetrouwbaarheid – zeker voor het toekennen van eigenschappen – wijst erop dat literaire evaluaties zich niet altijd eenduidig aan de hand van één objectief te bepalen set van categorieën laten beschrij-

105 Cláudio Rebelo De Sá en andere, "Mining association rules for label ranking", in *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (Springer, 2011), 433.

106 Karsdorp, *Retelling Stories. A Computational-Evolutionary Perspective*, 36.

107 Vgl. JHW Boswijk, "Er zal altijd wel iets over geschreven worden. De ontwikkelingen binnen de evaluaties in recensies aan de hand van het prozawerk van Remco Campert" (Masterscriptie, Radboud Universiteit, 2017), 46; JM Ter Meulen, "Een krant die nooit z'n ziel verkocht" (Masterscriptie, Radboud Universiteit, 2018), 56.

108 Sven Vitse, "Esther op de Beek, Een literair fenomeen van de eerste orde. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: Een kwantitatieve en kwalitatieve analyse.", *Tijdschrift voor tijdschriftstudies* 36 (2014): 149; Gerard Verhoeven, "Bespreking van Linders, Y. (2014). Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005. Proefschrift, Radboud Universiteit Nijmegen. 288 pp", *Tijdschrift voor taalbeheersing* 37 (2015): 105.

ven.[109] Een hypothetische verwijzing naar de verschillende verhaallijnen van een roman, bijvoorbeeld, kan door de ene codeur gelezen worden als een structuuraspect, terwijl deze in de ogen van een andere codeur juist een plotaspect markeert.[110] Voor de codes *stijl* en *dialogen*, voor *geheel/deel* en *overig*, voor *samenhang* en *continuïteit*, en voor *intensiteit* en *emotionaliteit* geldt een gelijksoortige ambiguïteit, zoals respectievelijk Verhoeven en Vitse hebben beargumenteerd.[111] In dergelijke gevallen is er geen objectief vast te stellen 'gouden standaard', geen ergens te achterhalen waarheid die ons vertelt dat de hypothetische verwijzing *ofwel* een plot-code, *ofwel* een aspect-code moet krijgen. In die zin is het ook onzinnig om methodologisch te pretenderen dat het gecodeerde recensiecorpus van Linders en Op de Beek die wel bevat. Maar wat we wél kunnen zeggen met een aanzienlijke zekerheid, is dat zowel een plotcode als een aspectcode in deze context *relevanter* zijn dan de dialogencode. En dat principe is precies hetgeen we via een multi-label ranking kunnen benaderen.

De tweede reden om voor een multi-label rankingbenadering te kiezen ligt sterk in het verlengde van de eerste reden, al heeft deze meer betrekking op een specifiek methodologische keuze van Linders en Op de Beek. Zoals Op de Beek aangeeft, hebben beide onderzoekers voor alle 734 recensies individueel evaluatieve uitspraken gemarkeerd, waarbij elke uitspraak die 'tenminste door één van de beide onderzoekers germarkeerd werd [...] van een evaluatiedomeincode voorzien' is.[112] Hoewel deze keuze, zoals Vitse terecht heeft opgemerkt, verder niet wordt toegelicht, heeft ze wel vergaande methodologische consequenties: het betekent namelijk dat de dataset coderingen bevat die nadrukkelijk *niet* op intersubjectieve wijze tot stand zijn gekomen, maar dus slechts door één van de twee onderzoekers is gecodeerd.[113] Daarmee geldt hier dan ook hetzelfde principe als bij de ambigue codes: dat één codeur een bepaalde passage heeft gecodeerd, betekent niet noodzakelijk dat hierover overeenstemming met de andere codeur is. Maar wederom kunnen we wel zeggen dat deze code in ieder geval wel relevanter zijn dan de codes die niet aan deze passages zijn toegevoegd – de onderzoeker kent deze code niet voor niets aan de passage toe.

109 Verhoeven, "Bespreking van Linders, Y. (2014). Met waardering gelezen. Een nieuw analyseinstrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005. Proefschrift, Radboud Universiteit Nijmegen. 288 pp", 104.

110 Of beide, natuurlijk.

111 Verhoeven, "Bespreking van Linders, Y. (2014). Met waardering gelezen. Een nieuw analyseinstrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005. Proefschrift, Radboud Universiteit Nijmegen. 288 pp", 104; Vitse, "Esther op de Beek, Een literair fenomeen van de eerste orde. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: Een kwantitatieve en kwalitatieve analyse.", 149.

112 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 127.

113 Een meer logische keuze zou mijns inziens zijn om eventuele conflictgevallen te bespreken met een derde codeur, zoals in Henk Pander Maat en andere, "Tekstkenmerken en tekstkwaliteit van leerlingteksten. Een annotatiestudie", *Tijdschrift voor Taalbeheersing* 41, nummer 2 (2019): 342.

4.3 FEATURES

Een belangrijke voorwaarde voor elk automatisch classificatiesysteem, voor elk *machine learning*-model, is dat we onze teksten numeriek kunnen representeren. Niet noodzakelijk als enen en nullen – het cliché waarmee oppervlakkige kritiek de inzet van digitale technologieën voor het lezen van teksten onterecht als reductionistisch tracht af te schilderen – maar als betekenisvolle reeksen van cijfers, ook wel vectorrepresentaties genoemd. Er zijn verschillende manieren om woorden en teksten te vectoriseren, variërend van heel simpel tot vrij complex, en er is geen manier die voor elk probleem, voor elke dataset het beste werkt.[114] In dit onderzoek vergelijk ik drie manieren van vectoriseren. Op oplopende volgorde van complexiteit zijn dat: 1) een *bag-of-words* model 2) *word embeddings*, en 3) *deep contextual word embeddings*. In wat volgt zal ik nader op elk van deze drie manieren ingaan.

4.3.1 *Bag-of-words*

Een van de meest simpele doch bijzonder effectieve methodes van numerieke tekstrepresentatie is het zogeheten *bag-of-words* (BoW) model. Dit model maakt allereerst een geordend overzicht van alle unieke woorden in het corpus. Dit noemen we het vocabulaire V, en de lengte van het vocabulaire (het aantal unieke woorden in het corpus) duiden we formeel aan met $|V|$. In het standaard *bag-of-words* model wordt elk document, in ons geval: elke zin, in het corpus gerepresenteerd door een vector $d = (q_1, ..., q_{|V|})$, waarbij q_j de frequentie van het j-de woord van V in het document d representeert. Minder formeel: we representeren elke zin als een lijst van frequenties, waarbij we voor elk uniek woord in het corpus noteren hoe vaak dit woord voorkomt in die specifieke zin, hetgeen in de praktijk vaak resulteert in een lijst die voor het grootste deel bestaat uit nullen – een zin telt immers slechts een beperkt aantal woorden van het totaal.

Ondanks dat het BoW-model voor veel tekstclassificatieproblemen bijzonder efficiënt en effectief werkt, kleven er aan de simpliciteit van het model ook enkele nadelen, die vrijwel allemaal te maken hebben met het feit dat er geen onderliggend begrip van taal zit ingebakken, maar slechts de aan- of afwezigheid van bepaalde reeksen van letters gescheiden door een witruimte in documenten aangeeft. Zo wordt woordvolgorde in de documenten genegeerd en beschouwt het twee vervoegingen van hetzelfde werkwoord formeel als twee totaal verschillende woorden. Ook kunnen algoritmes op basis van een BoW-model alleen maar patronen leren aan de hand van de woorden die in de trainingsset voorkomen. Wanneer in een ongezien document het woord 'stijltechnisch' voorkomt, maar dat woord nog niet is voorgekomen in de trainingsdata, zal het model het ongeziene document niet kunnen relateren aan documenten die het woord 'stijl' bevatten.

114 Ook wel de "No free lunch"theorem.

4.3.2 *Word embeddings*

Hoewel er voor bovengenoemde problemen allerlei ad hoc oplossingen bestaan, bieden *word embeddings* in veel gevallen een structurele uitkomst. *Word embeddings* zijn multidimensionale numerieke representaties van woorden waarbij woorden die gerelateerd zijn aan elkaar een gelijkaardige numerieke representatie hebben. Hierbij wordt een woord w dus niet meer louter en alleen gerepresenteerd aan de hand van zijn positie in een bepaald vocabulaire V, zoals in het *bag-of-word*s model, maar als een continue N-dimensionale vector $w = (x_1, ..., x_N)$, als een lijst van reëelwaardige getallen, waarvan elk getal een dimensie van de syntactische en semantische functie van dat woord dient te representeren.[115]

Modellen die *word embeddings* produceren zijn simpele, ongecontroleerde neurale netwerken die worden getraind op grote hoeveelheden aan teksten. Daarbij is het doel van zo'n netwerk om ofwel aan de hand van een gegeven context een woord te voorspellen (het CBOW NNLM), ofwel aan de hand van een gegeven woord de context te voorspellen (Skip-gram NNLM).[116] In beide gevallen is de achterliggende veronderstelling dat woorden die in gelijkaardige tekstuele context voorkomen, ook een gelijkaardige betekenis hebben.

Zoals Andrew Piper heeft beargumenteerd, is deze manier van computationeel modelleren van woordbetekenis 'in many ways a profound manifestation of post-structural textual theory' – 'however counterintuitve it may seem'.[117] De individuele getallen in de vectorrepresentatie van een bepaald woord hebben afgezonderd van andere woorden, van andere vectorrepresentaties, geen inherente, vaststaande betekenis; deze getallen krijgen pas betekenis in relatie tot andere woorden, wanneer er woorden in een ruimtelijke zin zich dichtbij of juist ver weg van dit bepaalde woord bevinden.[118] De betekenis van een woord is dus niet gefixeerd, maar afhankelijk van zijn relaties tot andere woorden, tot andere contexten.[119] En ook deze relaties zijn niet stabiel (vandaar nadrukkelijk post-structuralistische, en niet structuralistische of Saussuriaanse theorie), maar 'entirely contingent upon the scale of analysis'.[120] Ofwel: een *word embeddings*-model getraind op negentiende eeuwse romans zal heel andere relaties tussen dezelfde woorden onthullen dan een model getraind op twintigste eeuwse romans.

Eerder onderzoek heeft aangetoond dat de ogenschijnlijk triviale keuzes die gepaard gaan met het trainen van *word embeddings* kunnen leiden tot significantie verbetering (of verslechtering) van de presentaties van het uit-

115 Tomas Mikolov en andere, "Efficient estimation of word representations in vector space", *arXiv preprint arXiv:1301.3781*, 2013,

116 Deze context is doorgaans niet groter dan tien woorden aan linkerkant en tien woorden aan de rechterkant.

117 Piper, *Enumerations: Data and literary study*, 13.

118 Onder andere Jurafsky en Martin hebben evenzeer beargumenteerd dat deze manier van modelleren bij uitstek aansluit bij de ideeën van taalkundigen in de jaren '50, zoals Firth en Wittgenstein. Zie: Daniel Jurafsky en James H. Martin, "Speech and Language Processing" (Draftversie van 2 oktober 2019, beschikbaar via `https://web.stanford.edu/~jurafsky/slp3/`, 2019), hoofdstuk 6, 1-6.

119 Piper, *Enumerations: Data and literary study*, 14.

120 Ibidem, 13.

eindelijke classificatiemodel.[121] Waar het onderzoek eindigt met enkele nuttige aanbevelingen voor deze 'triviale' keuzes, borduur ik in deze scriptie voort op het vergelijkende perspectief door mij te richten op de keuze voor het corpus waarop *word embeddings* getraind worden, waarbij ik specifiek geïnteresseerd ben in de afweging tussen de grootte van het corpus en de domeinspecificiteit. Want hoewel corpusgrootte traditioneel als een belangrijke factor voor de prestaties van het model wordt aangewezen,[122] heeft recent onderzoek óók gewezen het op belang van domeinspecifieke, op relatief beperkte corpora getrainde *word embeddings*, bijvoorbeeld voor de classificatie van patenten, of jargonbeladen documenten in technische sector.[123] Dagbladrecensies vormen hierbij een interessante casus, omdat critici zich in het evalueren van de verschillende aspecten van boeken onvermijdelijk van een specifiek literairkritisch jargon bedienen dat tegelijkertijd – vanwege de specifieke plaatsing in het dagblad – ook begrijpelijk voor een breed publiek moet zijn.

In dit onderzoek vergelijk ik dan ook twee *word embeddings*-modellen. Het eerste is afkomstig van het CLiPS-onderzoekscentrum van de Universiteit van Antwerpen en is getraind op het Nederlandse COW (COrpora from the Web) corpus.[124] Het COW-corpus telt ruim 4 miljard woorden en bestaat uit teksten afkomstig van Nederlandse en Vlaamse websites.[125] Het tweede model heb ik zelf op getraind op 72.077 literatuurrecensies afkomstig van *LiteRom* (hierna: LitR),[126] waarbij ik gebruik heb gemaakt van de `Word2vec`-implementatie van *gensim*.[127] De *embeddings* zijn getraind met de eerdergenoemde CBOW-architectuur en kennen, evenals de voorgetrainde COW-*embeddings*, een dimensie van 320. Hoewel het LitR-corpus een substantieel deel van de Nederlandse literaire recensies die sinds 1945 gepubliceerd zijn bevat, is het aanzienlijk kleiner dan het veel diversere COW-corpus: het LitR-corpus telt ruim 67 miljoen woordtokens, verdeeld over ongeveer 3,5 miljoen zinnen. Daarmee is het COW-corpus ongeveer 60 keer groter dan het LitR-corpus.

Toch betekent dat niet noodzakelijk dat COW-*embeddings* tot betere resultaten zullen leiden voor deze classificatietaak dan de LitR-*embeddings*. Mijn hypothese is dat de *embeddings* die getraind zijn op het LitR-corpus niet alleen beter in staat zijn specifiek literairkritisch jargon te 'begrijpen', maar

121 Bhuwan Dhingra en andere, "A comparative study of word embeddings for reading comprehension", *arXiv preprint arXiv:1703.00993*, 2017,

122 Mikolov en andere, "Efficient estimation of word representations in vector space", 6.

123 Farhad Nooralahzadeh en andere, "Evaluation of Domain-specific Word Embeddings using Knowledge Resources", in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)* (2018); Julian Risch en Ralf Krestel, "Domain-specific word embeddings for patent classification", *Data Technologies and Applications* 53, nummer 1 (2019): 108–122.

124 Stéphan Tulkens en andere, "Evaluating unsupervised Dutch word embeddings as a linguistic resource", *arXiv preprint arXiv:1607.00225*, 2016, Embeddings zijn beschikbaar via `https://github.com/clips/dutchembeddings`.

125 Roland Schäfer en Felix Bildhauer, "Building large corpora from the web using a new efficient tool chain.", in *LREC* (2012), 486–493.

126 Zie hoofdstuk 5.1 voor een uitgebreide toelichting van de totstandkoming van dat corpus.

127 Radim Rehurek en Petr Sojka, "Software framework for topic modelling with large corpora", in *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (Citeseer, 2010).

ook een beter grip hebben op de situationeel bepaalde lexicale keuzes, om met Douglas Biber te spreken: het *register*, van literaire recensies.[128]

Laat me dit – wellicht nog enigszins abstracte – idee illustreren met een voorbeeld, waarvoor ik eerst nog een stapje terug zal nemen. Zoals ik eerder heb benoemd, is het centrale principe achter *word embeddings* dat woorden die een gelijkaardige semantische of syntactische functie hebben ook een gelijkaardige numerieke representatie hebben. Een van de meest gebruikelijke maatstaven om uit te drukken in hoeverre twee woordvectoren $\mathbf{w_i}$ en $\mathbf{w_j}$ een gelijkaardige numerieke representatie hebben, is de zogeheten cosinusgelijkenis tussen twee vectoren. Met de cosinusgelijkenismaatstaf berekenen we de cosinus van de hoek tussen twee vectoren:

$$\text{cosine}(\mathbf{w_i}, \mathbf{w_j}) = \frac{\sum_{k=1}^{n} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{n} w_{ik}^2} \sqrt{\sum_{k=1}^{n} w_{jk}^2}} \tag{1}$$

En dit neemt een waarde aan tussen -1 en 1, waarbij geldt dat hoe dichter de cosinusgelijkeniswaarde in de buurt van 1 komt, hoe groter de gelijkenis tussen de twee woordvectoren is. Zodoende kunnen we, gegeven een bepaald woord w_i, alle overige woorden in het vocabulaire sorteren op de cosinusgelijkheidswaarde met betrekking tot w_i, om daarmee vervolgens een top N meest vergelijkbare woorden te construeren. En zoals we in Tabel 4.2 zien, kunnen deze meest vergelijkbare woorden tussen de twee modellen onderling behoorlijk verschillen. We zien de top vijf meest vergelijkbare woorden voor *stijl*, *inhoud*, en *structuur* – drie woorden die weliswaar kunnen verwijzen naar een specifiek *aspect* van een literair werk maar tegelijkertijd, buiten de context van literaire kritiek, ook een bredere inzetbaarheid hebben. Het verschil tussen de twee modellen manifesteert zich dan ook precies langs deze dimensie. Waar *stijl* in het model getraind op het LitR-corpus uitsluitend wordt geassocieerd met woorden die betrekking hebben op vertellen en schrijven, zien we in het COW-model met 'bouwstijl' en de ietwat curieuze woorden 'vormentaal' en 'shaker-stijl' woorden die eerder verwant zijn aan bouwkunst en interieur. Ook *inhoud* krijgt in het LitR-model een specifiek literair-kritische invulling, terwijl we in het COW-model met drie websitenamen als meest vergelijkbare woorden vooral de ruis kunnen terugzien die het louter trainen op websiteteksten blijkbaar oplevert. Bij de meest vergelijkbare woorden met *structuur* zien we, tenslotte, de minste verschillen tussen de twee modellen – daar past alleen 'basistructuur' minder in een literair-kritisch register.

Ik wil nog een stap verder gaan. Omdat naast 'alledaagse' woorden zoals in Tabel 4.2 ook eigennamen een unieke numerieke representatie krijgen, hebben *word embeddings*-modellen mijns inziens de potentie om subtiele literairhistorische verhoudingen te incorporeren. De manier waarop de woordvectoren van de achternamen van twee auteurs zich bijvoorbeeld tot elkaar verhouden, wordt immers primair bepaald door de mate waarin twee auteurs in gelijkaardige tekstuele contexten voorkomen, door de mate

128 Douglas Biber en Susan Conrad, *Register, Genre, and Style* (Cambridge University Press, 2009), 9.

Tabel 4.2: Meest vergelijkbare woorden voor *stijl*, *inhoud* in *structuur*

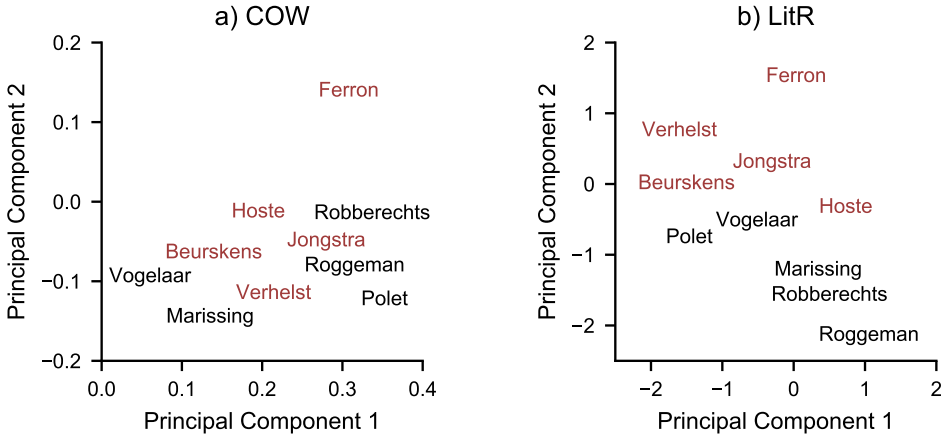
Corpus	Meest vergelijkbaar		
	Stijl	*Inhoud*	*Structuur*
LitR	schrijfstijl	struktuur	struktuur
	verteltrant	thematiek	opbouw
	schrijftrant	structuur	constructie
	vertelstijl	strekking	konstruktie
	schrijfwijze	vormgeving	compositie
COW	bouwstijl	hoedoen.be	struktuur
	vormentaal	leesvriendelijke	structuren
	stijlelementen	hurghada-info-site	basisstructuur
	shaker-stijl	presentatieblog.nl	stuctuur
	vormtaal	inschuifhoogten	gelaagdheid

waarin zij contextueel inwisselbaar zijn. En aangezien deze contextuele positionering van eigennamen in recensies – van vermeldingen van auteurs, van personages, van literaire stromingen – zeer vermoedelijk ook veel vaker op enigerlei wijze *literair* van aard is dan in willekeurige websiteteksten, verwacht ik dat de onderliggende literair-historische 'kennis' in het LitR-model aanzienlijk sterker is dan in het COW-model, waar de waarschijnlijkheid dat een naam zoals 'Claus' verwijst naar de Nederlandse prins in plaats van de Vlaamse auteur al veel groter wordt.

Om deze vermoedens enigszins te concretiseren, heb ik gekeken in hoeverre auteurs van twee literaire stromingen alleen op basis van de woordvectoren van hun achternamen van elkaar te onderscheiden zijn. Daarvoor heb ik een casus genomen waarvan het onderscheid voor veel onderzoekers wellicht niet direct hapklaar is (zeg: het onderscheid tussen de Vijftigers en de Maximalen), maar waarbij beide literaire bewegingen qua periodisering, poëtica, en receptie juist veel gelijkenissen vertonen, namelijk het onderscheid tussen het 'ander' proza en het postmodern proza. Figuur 4.3 toont het resultaat van een zogeheten *Principal Component Analysis*[129] (PCA) op de woordvectoren van de achternamen van vijf auteurs van het postmodern proza en vijf auteurs van het 'ander' proza. Ik ben daarbij uitgegaan van het onderscheid dat Sven Vitse in zijn proefschrift tussen beide varianten van experimenteel proza maakt.[130] Zoals we in de figuur kunnen zien, lopen de achternamen van de auteurs in het COW-model behoorlijk door elkaar heen. In het LitR-model, daarentegen, kunnen we twee duidelijke clusters van elkaar onderscheiden. Aan de bovenzijde zien we de auteurs van het postmodern proza, en aan de onderzijde die van het 'ander' proza. An-

129 Het doel van een *Principal Component Analysis* is om hoog-dimensionale data te reduceren naar enkele 'hoofdcomponenten' (een beperkt aantal variabelen), die zo veel mogelijk van de variantie binnen een bepaalde dataset verklaren. In dit geval gebruik ik PCA vooral om de *word embeddings* ruimtelijk te kunnen visualiseren. De *embeddings* die ik gebruik kennen een dimensie van 320, en zijn daarmee onmogelijk te visualiseren. Via een PCA kan ik de dimensies van mijn *embeddings* reduceren tot twee variabelen die tegelijkertijd zo veel mogelijk van de variantie in mijn originele dataset ondervangen.

130 Sven Vitse, *Een woestijn die de stad verpulvert. Marxisme, kritiek en representatie in poststructuralisme, Ander Proza en postmodern proza*, Vrije Universiteit Brussel, 2008.



Figuur 4.3: De twee *word embeddings*-modellen vergeleken via een tweedimensionale *Principal Component Analysis* op de achternamen van enkele auteurs van het postmodern proza (rood) en het 'ander' proza (zwart).

ders geformuleerd: in het LitR-model zijn de namen van de auteurs van het 'ander' proza en van het postmodern proza perfect lineair onderscheidbaar op de tweede PCA-dimensie.[131] Interessant is ook de plaatsing van Willy Roggeman, die voor Vitse 'de spreekwoordelijke vreemde eend in de bijt' vormt omdat hij 'een stem [laat] horen die volstrekt anders klinkt dan die van Vogelaar en co', en ook hier het verst van het centrum is verwijderd.[132] Tegelijkertijd moeten we oppassen niet te veel in de individuele plaatsing te lezen: zo vertoont het werk van Ferron volgens Vitse, in tegenstelling tot Verhelst, nog 'enige ideologische raakvlakken' met het 'ander' proza, maar is hij in deze figuur van alle postmoderne auteurs juist het verst van het 'ander' proza verwijderd.[133] De specifieke plaatsing van de auteurs derhalve daargelaten, geeft Figuur 4.3 ons desalniettemin een behoorlijke indicatie van het feit dat het LitR-model in staat is om subtiele literair-historische verhoudingen te reproduceren.

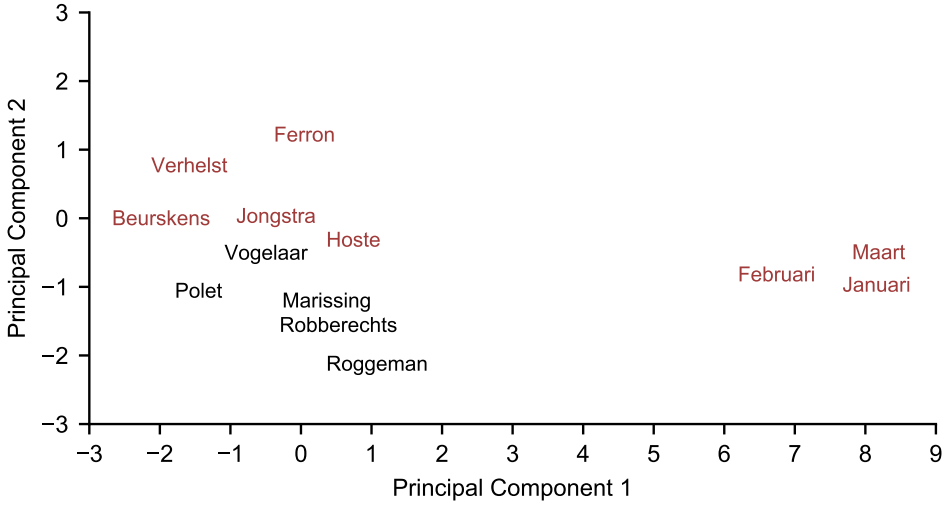
4.3.3 *Deep contextualized word embeddings*

Word embeddings zijn, in vergelijking met bag-of-words-modellen, een behoorlijke stap vooruit richting een daadwerkelijk computationeel begrip van taal. Toch zijn ook *word embeddings*-modellen niet vrij van reductionistische veronderstellingen. Een van de grootste problemen is het feit dat alle woorden in het vocabulaire, gegeven een bepaald corpus, maar één gefixeerde vectorrepresentatie hebben. We trainen de woordvectoren aan de hand van de context, maar veronderstellen vervolgens wel dat elke reeks van letters gescheiden door een witruimte, elk lexeem, altijd dezelfde betekenis heeft. Om te illustreren waarom deze veronderstelling problematisch kan zijn, voegen we aan de auteurs in Figuur 4.3 een postmoderne auteur toe die Vitse

131 Voor de kritische lezer: in het corpus waarop de *embeddings* zijn getraind zat geen recensie van Vitse's proefschrift of een (expliciete) verwijzing daarnaar.

132 Vitse, *Een woestijn die de stad verpulvert. Marxisme, kritiek en representatie in poststructuralisme, Ander Proza en postmodern proza*, 33.

133 Ibidem, 34.



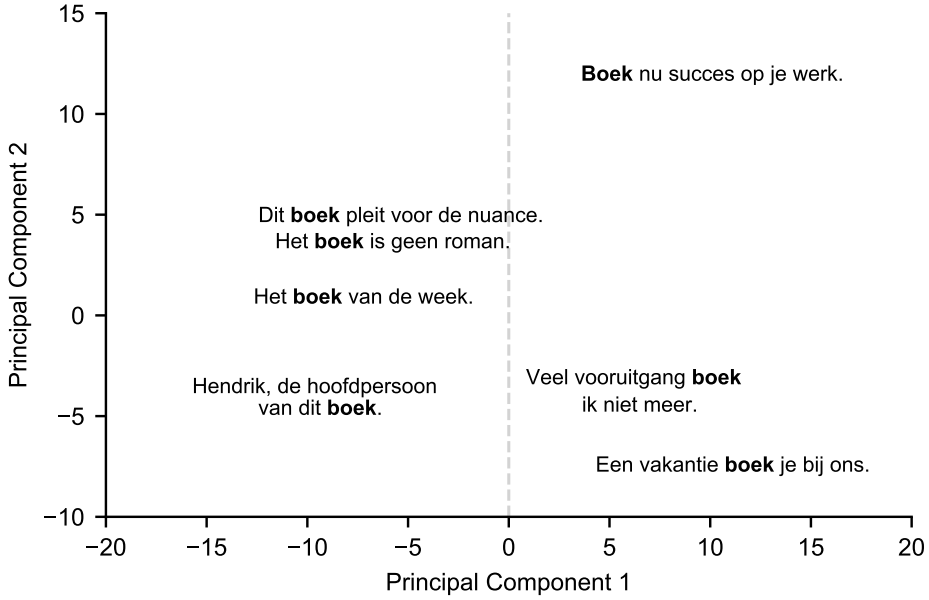
Figuur 4.4: Een tweedimensionale PCA op de LitR-embeddings van de achternamen van vijf auteurs van het postmodern proza en het 'ander' proza, en drie maanden.

in zijn proefschrift niet bespreekt: M. Februari. En laten we naast Februari ook de hyper-postmoderne want non-existente auteurs genaamd 'Januari' en 'Maart' aan de figuur toevoegen. Het resultaat: Figuur 4.4. *Februari*, zo blijkt uit deze figuur, wordt hier niet gezien als de auteur van postmodern proza, of überhaupt als een auteur, maar wordt, aanzienlijk ver verwijderd van de andere achternamen, juist geclusterd met maandaanduidingen. Dit betekent dat wanneer we een zin lezen als 'Februari schrijft scherp', de computer deze verwijzing naar *Februari* eerder zal 'interpreteren' als een tijdsaanduiding dan als naamaanduiding. En dit principe geldt niet alleen voor ambigue auteursachternamen, maar voor *alle* homonieme en polyseme woorden: voor *arm*, voor *boeken*, voor *punt* – enzovoorts.

De meest recente taalmodellen, zoals ELMo, BERT, of XLNet zijn dan ook gebaseerd op het idee van *deep contextualized word embeddings*.[134] Zoals de naam al enigszins verklapt, worden woorden hierbij niet meer gerepresenteerd door een gefixeerde vector, maar is de vectorrepresentatie van een woord volledig afhankelijk van de tekstuele context waarin dat woord voorkomt. Dit ondervangt niet alleen het probleem van homonieme en polyseme woorden, maar maakt het ook mogelijk om registergerelateerde verschillen, zoals we die bijvoorbeeld zagen in Tabel 4.2, te herkennen zonder daarvoor het model op een specifiek corpus van recensies te moeten trainen.

Dat laatste – het trainen van een eigen model – zou ook nauwelijks mogelijk zijn. De taalmodellen die *deep contexalized word embeddings* produceren zijn gigantische neurale netwerken met gemiddeld meer dan 100 miljoen parameters, die dagenlang worden getraind op meerdere krachtige pro-

134 Vgl. Matthew E Peters en andere, "Deep contextualized word representations", *arXiv preprint arXiv:1802.05365*, 2018, Devlin en andere, "Bert: Pre-training of deep bidirectional transformers for language understanding"; Yang en andere, "XLNet: Generalized Autoregressive Pretraining for Language Understanding".



Figuur 4.5: Een tweedimensionale PCA op de BERT-embeddings van het woord 'boek' in verschillende tekstuele contexten. De *embeddings* zijn geproduceerd met behulp van het Nederlandstalige model.

cessors.[135] Ik vergelijk in deze scriptie daarom twee voorgetrainde BERT-modellen. Het eerste is het meertalige BERT-model dat ontwikkeld is door de auteurs van het oorspronkelijke artikel waarin BERT gepresenteerd is. Dit model is getraind op *alle* Wikipedia-teksten van 104 talen, waaronder het Nederlands.[136] Het tweede model is een specifiek Nederlands BERT-model (*BERTje*), recentelijk ontwikkeld door een groep onderzoekers aan de Rijksuniversiteit Groningen, en getraind op een meer diverse dataset dan het meertalige model, bestaande uit onder andere romans, nieuwsberichten en Wikipedia-artikelen.[137]

BERT wordt getraind door willekeurige *tokens* in een tekst weg te laten, waarbij BERT aan de hand van de linker- en rechtercontext het weggelaten woord moet voorspellen – vergelijkbaar met zogeheten cloze-toetsen voor tekstbegrip.[138] Ter illustratie van de werking van contextueel-bepaalde *word embeddings* toont Figuur 4.5 het resultaat van een PCA op de BERT-*embeddings* van het woord 'boek' in verschillende tekstuele contexten. Om de BERT-*embeddings* te generen heb ik gebruik gemaakt van het Nederlandstalige BERT-model via `bert-as-service`.[139] In de voorbeeldzinnen kunnen

135 Journalist Tony Peng schat de kosten van het trainen van XLNet, op basis van de computationele specificaties in het originele paper, in op $61,000, terwijl het trainen van het grootste BERT-model ongeveer $6,912 zou kosten. Het lijkt me vanzelfsprekend dat ik dergelijke middelen voor deze masterscriptie niet heb. Zie: Tony Peng, "The Staggering Cost of Training SOTA AI Models", *Synced*, juni 2019, https://syncedreview.com/2019/06/27/the-staggering-cost-of-training-sota-ai-models/.

136 Zie ook Telmo Pires en andere, "How multilingual is Multilingual BERT?", *arXiv preprint arXiv:1906.01502*, 2019,

137 Wietse de Vries en andere, *BERTje: A Dutch BERT Model*, 2019, arXiv: 1912.09582 [cs.CL].

138 Devlin en andere, "Bert: Pre-training of deep bidirectional transformers for language understanding", 1-2.

139 Han Xiao, *bert-as-service*, https://github.com/hanxiao/bert-as-service, 2018.

we twee varianten van 'boek' onderscheiden: het enkelvoud van het werkwoord boeken, en het zelfstandig naamwoord boek. In de figuur kunnen we duidelijk zien hoe deze verschillen betekenissen worden onderscheiden: de vier zelfstandig naamwoordensvarianten zijn aan de linkerkant van de figuur geplaatst, terwijl de werkwoordsvarianten meer aan de rechterzijde geplaatst zijn – al lopen deze, vermoedelijk omdat de context van deze zinnen ook meer verschilt, wel meer uiteen.

4.4 MODELLEN

Met de vectorrepresentaties van de zinnen en de daarbij behorende labels (zoals besproken in paragraaf 4.1) hebben we nu twee van drie essentiële ingrediënten voor ons automatisch classificatiesysteem verzameld. Het laatste – en wellicht ook wel het meest belangrijke – ingrediënt van de taart der *machine learning* is een functie f om de *input* X om te zetten naar de gewenste output Y, ofwel: $f : X \rightarrow Y$. De daadwerkelijke *machine learning*-modellen kunnen we, vervolgens, dan ook beschouwen als een manier waarop we f kunnen benaderen.

Net als voor het verkrijgen van vectorrepresentaties van taal is er geen manier, geen *machine learning*-model dat voor elk voor elk probleem en voor elke dataset het beste werkt. Ik vergelijk daarom drie verschillende modellen. Wederom in oplopende volgorde van complexiteit zijn dat: 1) een logistische regressiemodel 2) een *bi-directional gated recurrent unit*-netwerk en 3) BERT met een classificatielaag. Anders dan in de vorige paragraaf zal ik geen uitgebreide uitleg geven van de werking van deze modellen – daarvoor zijn de modellen te complex en is de ruimte te kort. Ik ga wel kort in op het centrale idee achter deze modellen, maar zal vooral aandacht besteden aan de specifieke keuzes die ik bij de inzet van de modellen heb gemaakt.

4.4.1 Logistische regressie

Logistische regressie is vooral bekend als een statistisch analysemodel dat bijzonder populair is binnen de exacte en sociale wetenschappen, maar wordt in de *Natural Language Processing* ook gebruikt voor allerlei classificatietaken.[140] In een simpele, binaire logistische regressie generen we aan de hand van een inputvector x de waarschijnlijkheid dat x tot een bepaalde klasse (bijvoorbeeld: een evaluatieve uitspraak) behoort. Daarvoor leren we voor elk element x_i in x een bijbehorend gewicht w_i , in combinatie met een zogeheten *bias term* b. Gegeven een BOW-vectorrepresentatie van een zin, bijvoorbeeld, kunnen we verwachten dat een woord als 'ijzersterk' een hoog gewicht toegekend krijgt wanneer we evaluatieve uitspraken willen classificeren. De lineaire combinatie van w, x, en b ($w\dot{x} + b$) transformeren we vervolgens via de zogeheten *sigmoid*-functie naar een waarschijnlijk tussen 0 en 1.

140 Jurafsky en Martin, "Speech and Language Processing", 75.

In deze scriptie maak ik gebruik van de *scikit-learn*-implementatie van logistische regressie met L_2-regularisatie.[141] De regularisatiesterke C is gezet op 1. Voor de multi-label rankingproblemen gebruik ik de `OneVsRestClassifier`-wrapper in *scikit-learn* om logistische regressie geschikt te maken voor het voorspellen van meerdere labels. Hierbij wordt voor elke code in onze dataset (*geheel/deel*, *stijl*, *dialogen*, enzovoorts) een aparte logistische regressie-*classifier* getraind, die de (waarschijnlijkheid van de) aan- of afwezigheid van die specifieke code gegeven een bepaalde zin voorspelt. Dit staat ook wel bekend als de *binary relevance*-strategie. We kunnen vervolgens een ranking produceren door de codes op waarschijnlijkheid te sorteren.

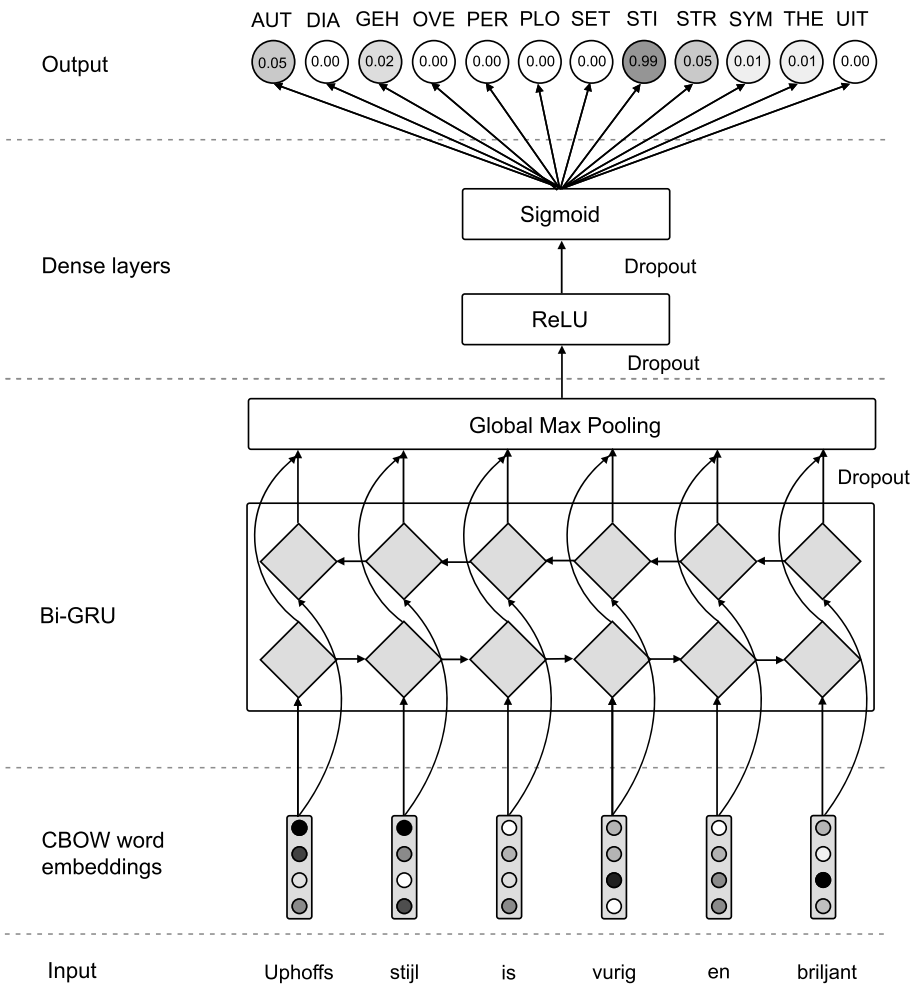
Voor elke output, voor elke waarschijnlijkheid tussen 0 en 1, verwacht een binair logistisch regressiemodel één gefixeerde N-dimensionale vector als *input*. Met andere woorden: wanneer we op zinsniveau classificeren, moeten we onze teksten ook op zinsniveau – en dus niet op teken-, woord- of recensieniveau – representeren. Dit is al het geval voor de BOW-representatie van ons corpus: daar representeren we elke zin als in een lijst van frequenties van woorden die in die zin voorkomen. Voor *word embeddings* geldt dat echter nog niet, omdat *elk* woord in een zin al gerepresenteerd wordt door een N-dimensionale vector. Een simpele oplossing voor dit probleem is om het gemiddelde van alle numerieke representaties van de woorden in een zin te berekenen. Dit noemen we *average pooling*, en dat is ook de methode die ik in deze scriptie hanteer. Een nadeel van deze methode is dat we hiermee de informatie over de volgorde van de woorden negeren.

4.4.2 Bi-GRU

Dat een standaard logistisch regressiemodel één gefixeerde N-dimensionale vector vereist, maakt het dan ook relatief moeilijk om informatie over de woordvolgorde goed te kunnen exploiteren. Dergelijke informatie is echter in veel gevallen wel betekenisvol: taal is immers, zoals Daniel Jurafsky en James H. Martin stellen, een 'inherently temporal phenomenon', en met het negeren van die temporaliteit, met het negeren van de woordvolgorde, gaat veel waardevolle informatie verloren.[142] Weliswaar zouden een logistische regressie-model kunnen toepassen op contextuele *word embeddings*, en zodoende alsnog sequentiële informatie verdisconteren in de voorspelling, maar idealiter hebben we een model waarbij we de gewichten voor de representatie en de classificatie tegelijkertijd kunnen afstemmen. Mede daarom maken veel van de recente toepassingen van *machine learning* op talige fenomenen gebruik van modellen die specifiek ontworpen zijn voor sequentiële data.

Een *Recurrent Neural Network* (RNN) is zo een variant van een neuraal netwerk dat specifiek bedoeld is voor sequentiële data. Hoewel neurale netwerken onmetelijk veel complexer zijn dan regressiemodellen, zijn ze wiskundig gelijkaardig aan logistische regressie: ook in neurale netwerken leren we gewichten voor een bepaalde *input*.[143] In een RNN verwerken we een zin

141 F. Pedregosa en andere, "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research* 12 (2011): 2825–2830.
142 Jurafsky en Martin, "Speech and Language Processing", 169.
143 Ibidem, 123.



Figuur 4.6: Een illustratie van de werking van het Bi-GRU-model, met het voorspellen van aspectcodes van de zin 'Uphoffs stijl is vurig en briljant' als voorbeeld. De zin zetten we eerst om in numerieke woordrepresentaties via onze voorgetrainde LiTR-embeddings. Deze numerieke representaties verwerken van links naar rechts en van rechts naar links in het netwerk, waarbij de code *stijl* (hier: STI) uiteindelijk de hoogste waarschijnlijkheid krijgt toegewezen (0.99).

woord voor woord, waarbij telkens het resultaat van de verwerking van het vorige woord wordt meegenomen bij de verwerking van het huidige woord. Door deze recursieve natuur zit er bij de verwerking van het laatste woord informatie over de hele zin ingebakken. Het resultaat van de verwerking van het laatste woord wordt dan ook veelal gebruikt om de uiteindelijke voorspelling te doen over de klasse van de zin.

Ik maak in deze scriptie gebruik van een geoptimaliseerde variant van een RNN, een zogeheten *bi-directional gated recurrent unit* (Bi-GRU).[144] Een GRU *onthoudt* niet alleen informatie over de verwerking van de voorgaande woorden, maar leert ook wanneer het informatie moet *vergeten*, waarmee het veel efficienter kan classificeren. De bi-directionale variant verwerkt daarbij

144 Kyunghyun Cho en andere, "Learning phrase representations using RNN encoder-decoder for statistical machine translation", *arXiv preprint arXiv:1406.1078*, 2014,

niet alleen woorden van links naar rechts, maar ook van rechts naar links. Voor de implementatie van mijn Bi-GRU-model gebruik ik *Keras*.[145]

Figuur 4.6 toont de specifieke architectuur van mijn model met het voorspellen van de aspectcodes op basis van LitR-embeddings als voorbeeld.[146] De *embeddings* zijn tijdens het trainen, enerzijds vanuit efficiëntie-oogpunt, en anderzijds vanuit het idee dat zeker de LitR-embeddings al voorgetraind zijn op domeinspecifieke teksten, tijdens het trainen gefixeerd. Omdat mijn dataset redelijk beperkt is, zeker in de context van het classificeren van aspect- en eigenschapscodes (ongeveer 7000 voorbeelden), kies ik voor een eenlaagse Bi-GRU. Daarbij gebruik ik *dropout* in drie lagen van het netwerk[147] Als *optimizer* gebruik ik het recentelijk voorgestelde *Rectified Adam* (RAdam) optimalisatiealgoritme, dat in experimentele fase van het onderzoek, zoals de auteurs van het RAdam-paper ook betogen,[148] aanzienlijk stabieler bleek te convergeren dan het traditionele Adam-optimalisatiealgoritme.[149] Ten slotte gebruik ik voor zowel het binaire classificatieprobleem als voor het ranken van de aspect- en eigenschapscodes een *sigmoid*-functie voor de *output*-laag met *Binary Cross Entropy Loss*, hetgeen voor de multi-label-problemen vergelijkbaar is met de hierboven genoemde *binary relevance*-strategie.

4.4.3 BERT

In de vorige sectie (4.3.3) presenteerde ik BERT als een manier om contextuele *word embeddings* te produceren. We zouden deze *embeddings* vervolgens kunnen gebruiken als *features* voor een ander classificatiemodel, zoals logistische regressie. Dit is wat de auteurs van het originele BERT-paper de 'feature-based approach' noemen.[150] In de praktijk worden de beste resultaten echter gehaald door BERT niet alleen in te zetten voor het produceren van tekstrepresentaties, maar ook voor het classificeren – de zogeheten 'fine-

145 François en andere Chollet, *Keras*, `https://keras.io`, 2015.

146 In de resultaten zal ik uiteraard ook de werking van het model op basis van COW-embeddings vergeleken.

147 Met *dropout* voorkomen we dat het netwerk die zinnen die we gebruiken om het netwerk te trainen min of meer letterlijk gaat onthouden, waardoor het netwerk niet kan generaliseren op nieuwe voorbeelden (*overfitting*. Zeker voor beperkte datasets vormt *dropout* een bijzonder effectieve maatregel. Voor de lagen van GRU gebruik ik zowel een dropout (van 0.15) op de input als een dropout (van 0.1) op de verwerking van de *recurrent state*, tussen de pooling-laag en de ReLU zit dropout van 0.3 en tussen de ReLU en de sigmoid-laag een beperkte dropout van 0.1. Voor de waardes en de plaatsing van de dropout-lagen heb ik handmatig verschillende configuraties vergeleken, en het model geselecteerd waarbij de verhouding tussen presentaties op de trainingsset en de validatieset het minst snel uit balans raakte zonder daar aan prestatie op de validatieset voor in te leveren. Zie ook: Nitish Srivastava en andere, "Dropout: a simple way to prevent neural networks from overfitting", *The journal of machine learning research* 15, nummer 1 (2014): 1929–1958.

148 Liyuan Liu en andere, "On the variance of the adaptive learning rate and beyond", *arXiv preprint arXiv:1908.03265*, 2019,

149 Dit is nadrukkelijk op basis van handmatige vergelijking – voor een structurele vergelijking is er in deze scriptie geen ruimte. Meer specifiek was mijn indruk dat RAdam niet noodzakelijk betere resultaten opleverde, maar dat meerdere runs van hetzelfde model wel – relatief onafhankelijk van de *learning rate* – meer dezelfde prestaties op de validatieset opleverde. Omdat RAdam verder op moment van schrijven nog niet officieel geïmplementeerd is in *Keras*, gebruik ik een onofficiële implementatie van `https://github.com/CyberZHG/keras-radam`.

150 Devlin en andere, "Bert: Pre-training of deep bidirectional transformers for language understanding", 2.

Tabel 4.3: Beschrijving van de dataset na het splitsen van de train-, test- en validatieset. $|Y|$ markeert het aantal labels, $|D_{set}|$ de grootte van een subset van D.

Classificatietaak	$\|Y\|$	$\|D_{train}\|$	$\|D_{test}\|$	$\|D_{val}\|$
Evaluatieve uitspraken	2	22392	2800	2799
Aspectcodes	12	6972	872	871
Eigenschapscodes	15	6972	872	871

tuning approach'.[151] Hierbij beginnen we met het originele, voorgetrainde BERT-model, en voegen daar een lineaire laag (vgl. logistische regressie) bovenop aan toe voor het classificeren. Vervolgens *fine-tunen* we het model op onze classificatietaken. Dit houdt in dat we, in tegenstelling tot bijvoorbeeld het Bi-GRU-model,[152] niet helemaal vanaf nul beginnen, maar de gewichten van het voorgetrainde model gebruiken en deze (slechts) bijstellen voor onze classificatietaak. Omdat de gewichten van het voorgetrainde model al getraind zijn miljoenen teksten, zit hier immers al een behoorlijk (contextueel) begrip van taal ingebakken, hetgeen we goed voor de classificatie kunnen gebruiken. Alleen de bovenste laag, de *output*-laag, moeten we volledig 'from scratch' leren.[153] Dit maakt dat het *fine-tunen* van een BERT-model in veel gevallen niet alleen veel efficienter is dan het trainen van een Bi-GRU, maar ook dat je in de praktijk veel minder taakspecifieke voorbeelden nodig hebt om een model van een dusdanige omvang accuraat te kunnen gebruiken.[154]

De 'fine tuning approach' is ook de aanpak die ik in deze scriptie hanteer. Ik *fine-tune* BERT voor het binaire classificatieprobleem – de classificatie van evaluatieve uitspraken – via *HuggingFace's Transformers*[155] met *Tensorflow 2.0*-integratie.[156] Omdat er voor het fine-tunen van BERT relatief veel computerkracht nodig is, draai ik mijn scripts in een *Google Colab*-omgeving,[157] waarin een gratis GPU ter beschikking wordt gesteld. Ik *fine-tune* het model voor maximaal tien *epochs*; één *epoch* kost ongeveer vier minuten. Voor de multi-label rankingproblemen gebruik ik *Simple Transformers*,[158] een Python-bibliotheek op basis van *Huggingface's Transformers* die specifiek het fine-tunen van BERT voor multi-label problemen ondersteunt. Wederom draai ik het model voor niet langer dan tien *epochs*, en een *epoch* kost hier ongeveer zes minuten in de *Colab*-omgeving.

151 Devlin en andere, "Bert: Pre-training of deep bidirectional transformers for language understanding", 2.
152 Al gebruiken we bij het Bi-GRU ook wel voorgetrainde *embeddings*.
153 Devlin en andere, "Bert: Pre-training of deep bidirectional transformers for language understanding", 14.
154 Ibidem, 5.
155 Thomas Wolf en andere, "HuggingFace's Transformers: State-of-the-art Natural Language Processing", *ArXiv* abs/1910.03771 (2019).
156 Martín Abadi en andere, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015, https://www.tensorflow.org/.
157 https://colab.research.google.com
158 https://github.com/ThilinaRajapakse/simpletransformerst

4.5 EVALUATIE

Voor de evaluatie van de verschillende numerieke tekstrepresentaties besproken in paragraaf 4.3 en de modellen besproken in paragraaf 4.4 deel ik de dataset op in drie delen: een trainingsset, een testset en een validatieset. 80% van de dataset is gereserveerd om de modellen te trainen. 10% gebruik ik om de modellen te testen, en de laatste 10% gebruik als validatieset om de neurale netwerken te optimaliseren, waarbij ik het model dat het beste scoort op de validatieset selecteer. Het precieze aantal voorbeelden na splitsing is te vinden in Tabel 4.3.

Ik meet en vergelijk de presentaties van de verschillende modellen aan de hand van een aantal numerieke maatstaven. In wat volgt ga ik kort op elk van deze maatstaven in. Omdat multi-label rankingproblemen op een fundamenteel andere manier geëvalueerd moeten worden dan binaire classificatieproblemen,[159] gebruik ik voor deze twee verschillende problemen ook andere maatstaven, die ik apart van elkaar zal bespreken. De prestaties van de hierboven gepresenteerde *features* en modellen zet ik daarnaast af tegen een *baseline*-model dat geen *features* gebruikt, maar alleen voorspellingen doet op basis van simpele heuristieken. Voor het binaire classificatieprobleem gebruik een *random* baseline; voor de multi-label rankingproblemen gebruik ik, omdat een *random* baseline daar een te pessimistisch beeld van de complexiteit van de taak schetst, een baseline-model dat de labels steeds op dezelfde wijze ordent aan de hand van de frequentie van de labels in de trainingsset.[160]

4.5.1 Maatstaven binaire classificatie

Bij een binair classificatieprobleem zijn er in essentie vier manieren waarop de voorspelling van de computer en het daadwerkelijke label (het menselijke oordeel) zich tot elkaar kunnen verhouden. Met het classificeren van evaluatieve uitspraken als voorbeeld, kan een computer ten eerste een zin terecht (in de zin van: overeenkomstig met het menselijke oordeel) als een evaluatieve uitspraak markeren (*true positive*). Ten tweede kan een computer een uitspraak onterecht als evaluatief markeren (*false positive*). Ten derde kan een computer een uitspraak terecht als niet-evaluatief markeren (*true negative*). Ten vierde kan een computer een uitspraak onterecht als niet-evaluatief markeren (*false negative*). De drie maatstaven die ik in deze scriptie voor het binaire classificatieprobleem hanteer – *precision*, *recall* en F_1 – zijn in essentie dan ook niets anders dan manieren waarop we deze verschillende verhoudingen tegen elkaar kunnen afwegen.

Precision geeft, allereerst, aan hoeveel van de zinnen die de computer als evaluatieve uitspraak geclassificeerd heeft, daadwerkelijk evaluatieve uitspraken zijn. Dit meten we door de *true positives* te delen door de som van *false positives* en *true positives*:

$$\text{Precision} = \frac{\text{true postives}}{\text{true positives} + \text{false postives}} \tag{2}$$

159 Grigorios Tsoumakas en andere, "Mining multi-label data", in *Data mining and knowledge discovery handbook* (Springer, 2009), 667.

160 De optie 'prior' in scikit-learn's `DummyClassifier`.

Omdat *precision* geen indicatie geeft van de evaluatieve uitspraken die de computer heeft gemist, gebruiken we *recall* om aan te geven hoeveel van alle evaluatieve uitspraken er door computer ook als zodanig zijn geclassificeerd. Dit meten we door de *true positives* te delen door de som van *false negatives* en *true positives*:

$$\text{Recall} = \frac{\text{true postives}}{\text{true positives} + \text{false negatives}} \tag{3}$$

De F_1-score, tenslotte, is het gewogen gemiddelde van *precision* en *recall*:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

Deze score gebruik ik dan ook de prestaties van de modellen op de validatieset te vergelijken,[161] en zal – omdat ze dus zowel *precision* als *recall* incorporeert – fungeren als mijn primaire maatstaf voor de uiteindelijke evaluatie van de modellen.

4.5.2 Maatstaven multi-label ranking

Waar we in de context van binaire classificatie nog kunnen spreken over fout en goed, over zuiver terecht en onterecht, geldt dat voor modellen die een ranking produceren niet. Immers: wanneer een van de twee relevante labels van een zin hoog in voorspelde ranking staat, maar de andere juist laag, is deze ranking dan goed of fout? De ranking is wellicht niet perfect, maar het antwoord ligt vermoedelijk ergens in het midden. Daarom gebruiken we voor multi-label rankingsproblemen alternatieve maatstaven die allemaal een indicatie geven van de mate waarin de relevante labels hoog in de ranking staan en vice versa, zij het met een andere invalshoek. Ik gebruik er vier: *Average precision*, *One Error*, *Is Error* en *Ranking loss*.[162]

Average Precision meet in hoeverre de relevante labels hoog in de ranking staan aan de hand van de gemiddelde fractie van labels die boven de relevante labels in de ranking staan. Laten we als voorbeeld een zin nemen die de codes *geheel/deel*, en *auteur/intentie* toegewezen heeft gegeven. De computer produceert aan de hand van deze zin een ranking, waarbij de code *geheel/deel* bovenaan gerangschikt is, en de code *auteur/intentie* als vierde. We meten de *precision* van een code door het aantal correcte labels tot en met de positie van de code te delen door de positie van de code. Voor *geheel/deel* is de *precision* dan $1/1 = 1$; voor *auteur/intentie* $2/4 = 0.5$. De *Average Precision* van één ranking is de gemiddelde *precision* van alle relevante codes. Wanneer we, gegeven een labelset $L = \{\lambda_1, ..., \lambda_c\}$, de ranking van een la-

161 Ik kies hier voor F_1, en niet voor bijvoorbeeld *loss*, omdat de modellen met de laagste *loss* vooral leken te resulteren in modellen die heel goed waren in het herkennen van niet-evaluatieve uitspraken, terwijl we idealiter een model hebben dat zowel niet-evaluatieve als evaluatieve uitspraken kan relatief accuraat kan herkennen.

162 Daarmee neem ik in grote mate de maatstaven die worden gebruikt door Karsdorp (2016) en Tsoumakas en andere (2009) over. Vgl. Karsdorp, *Retelling Stories. A Computational-Evolutionary Perspective*, 41; Tsoumakas en andere, "Mining multi-label data", 679-680.

bel λ noteren als $r_i(\lambda)$,[163] kunnen we de *Average Precision* over alle rankings formeel noteren als:

$$AP = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i|}\sum_{\lambda\in Y_i}\frac{|\{\lambda' \in Y_i : r_i(\lambda') \leqslant r_i(\lambda)\}|}{r_i(\lambda)} \tag{5}$$

En dit neemt een waarde aan tussen 0 en 1, waarbij 1 markeert dat alle labels in de dataset perfect geordend zijn.

Is Error, ten tweede, markeert het aantal rankings in de dataset dat *niet* perfect geordend is. Dit bereken we door het aantal rankings met een *Average Precision* lager dan 1 af te zetten tegen het totaal aantal rankings:[164]

$$\text{Is Error } = \frac{1}{n}\sum_{d=1}^{n} AP(d) < 1 \tag{6}$$

En ook dit neemt een waarde tussen 0 en 1, waarbij 0 markeert dat alle rankings perfect geordend zijn (ofwel: een *Average Precision* van 1 hebben).

One Error, ten derde, meet het aantal documenten waarbij het hoogst geplaatste label in de voorspelde ranking niet relevant is (dus: niet door de menselijke codeurs als zodanig is aangewezen).

$$\text{One Error} = \frac{1}{n}\sum_{i=1}^{m}\delta\left(\arg\min_{\lambda\in L} r_i(\lambda)\right) \tag{7}$$

Waarbij $\delta(\lambda)$ de waarde van 1 aanneemt wanneer het label niet relevant is, en 0 wanneer dit wel geval is, ofwel:

$$\delta(\lambda) = \begin{cases} 1 & \text{if } \lambda \notin Y_i \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

De *One Error*-maatstaf loopt van 0 tot 1. Een waarde van 0 geeft aan dat alle hoogstgeplaatste labels relevant zijn; een waarde van 1 dat alle hoogstgeplaatste labels irrelevant zijn.

Ranking loss, ten slotte, drukt het gemiddelde aantal keer dat irrelevante labels hoger geplaatst zijn dan relevante labels uit:[165]

$$RL = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i|\,|\bar{Y}_i|}\left|\left\{(\lambda_a,\lambda_b) : r_i(\lambda_a) > r_i(\lambda_b), (\lambda_a,\lambda_b) \in Y_i \times \bar{Y}_i\right\}\right| \tag{9}$$

Waarbij $\bar{Y}_i$ bestaat uit alle irrelevante labels (alle labels in L die niet in Y_i zitten). De *Ranking Loss* loopt eveneens van 0 tot 1, waarbij een score van 0 een perfecte ranking aangeeft.

4.6 RESULTATEN

4.6.1 Classificeren van evaluatieve uitspraken

Tabel 4.4 toont de resultaten voor het classificeren van evaluatieve en niet-evaluatieve uitspraken. Omdat de labels enigszins ongelijk verdeeld zijn

163 Ik neem hier, en in het vervolg van deze paragraaf, grotendeels de notatie over van: Tsoumakas en andere, "Mining multi-label data".
164 Karsdorp, *Retelling Stories. A Computational-Evolutionary Perspective*, 42.
165 Tsoumakas en andere, "Mining multi-label data", 681.

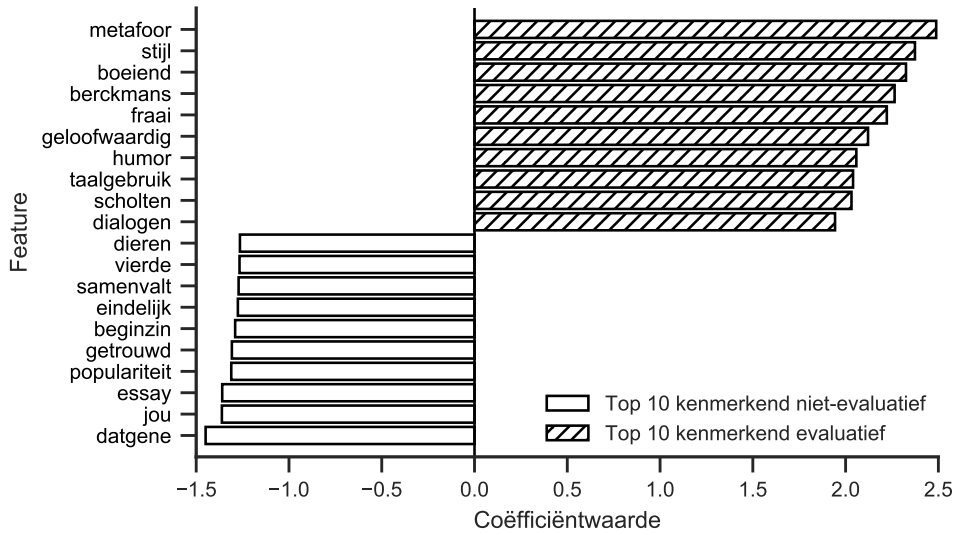
Tabel 4.4: Resultaten van de verschillende modellen en *features* voor het classificeren van evaluatieve uitspraken. Het model dat lichtrood gemarkeerd is, is het bestscorende model wanneer we kijken naar de gewogen F1-score.

		Niet-evaluatief			Evaluatief		
Model	Features	Precision	Recall	F1	Precision	Recall	F1
Random baseline		0.69	0.49	0.58	0.30	0.50	0.37
LR	BoW	0.81	0.89	0.85	0.68	0.51	0.58
	COW$_{\text{BIG}}$	0.77	0.92	0.84	0.68	0.38	0.49
	LitR	0.80	0.88	0.84	0.64	0.49	0.56
	BoW + LitR	0.82	0.88	0.85	0.66	0.55	0.60
	BoW + COW$_{\text{BIG}}$	0.81	0.89	0.85	0.67	0.53	0.59
Bi-GRU	LitR	0.87	0.86	0.86	0.68	0.70	0.69
	COW$_{\text{BIG}}$	0.88	0.82	0.86	0.64	0.74	0.68
BERT$_{\text{MULTILINGUAL}}$		0.84	0.87	0.85	0.70	0.66	0.68
BERT$_{\text{DUTCH}}$		0.86	0.89	0.87	0.75	0.69	0.72

(ongeveer 30% van de dataset bestaat uit evaluatieve uitspraken), geef ik de *precision*-, *recall*- en F_1-score (zoals besproken in paragraaf 4.5.1) voor het classificeren van niet-evaluatieve uitspraken en voor het classificeren van evaluatieve uitspraken apart weer. En zoals we in de tabel dan ook kunnen zien, zijn alle modellen ongeveer even goed in het classificeren van niet-evaluatieve uitspraken. Sommige modellen, zoals LR-COW$_{\text{BIG}}$, zijn wat voorzichtiger in het classificeren dan gemiddeld, maar over het algemeen loopt de F_1-score van 0.84 tot 0.87.

Als we daarentegen alleen kijken naar de classificatie van evaluatieve uitspraken, zien we al veel grotere verschillen. De neurale netwerken scoren aanmerkelijk beter dan de regressiemodellen, met een verschil van ongeveer 10 procentpunt op de F_1-score. Van de neurale modellen haalt het Nederlandse BERT-model veruit de beste resultaten. Dit model scoort ongeveer 3 procentpunt hoger dan de andere neurale modellen. De prestaties van het meertalige BERT-model zijn daarentegen niet significant beter dan die van de veel simpelere GRU-modellen, hetgeen een afweging suggereert tussen de complexiteit van het model en de specificiteit van de trainingsdata, waarin het Nederlandse BERT-model de beste balans in weet te vinden.

Ook de verschillende manieren van vectoriseren leveren interessante onderlinge prestatieverschillen op. Allereerst lijken deze resultaten de in paragraaf 4.3.2 uiteengezette hypothese dat *word embeddings* getraind op een beperkt domeinspecfiek corpus informatiever zijn dan *embeddings* die getraind zijn op een veel groter maar algemener corpus te bevestigen. Dit geldt vooral bij het LR-model; bij het Bi-GRU-model zijn de verschillen al veel kleiner. Eveneens opmerkelijk is het feit dat het LR-model op basis van de veel simpelere BoW-*features* beter scoort dan de modellen op basis van *word embeddings*, al werkt een combinatie van beide het best. Omdat de *embeddings* echter wel tot betere resultaten leiden in het Bi-GRU-model, kan de verklaring hiervoor vooral gevonden worden in de noodzaak om de *word embeddings* naar zinsrepresentaties te transformeren door het gemiddelde over alle woorden te nemen, waarbij onbelangrijke woorden een even grote invloed op het gemiddelde hebben als belangrijke woorden.



Figuur 4.7: De top 10 woorden met de hoogste (kenmerkend voor evaluatieve uitspraken) en laagste (kenmerkend voor niet-evaluatieve uitspraken) coëfficiëntwaarde in het logistische regressiemodel op basis van de *bag-of-words*-representaties. Deze woorden stuurden het model het sterkst richting één van de twee labels.

Een voordeel van het simpele logistische regressiemodel op basis van BoW is dat we heel helder kunnen traceren welke woorden een grote invloed uitoefenen op de uiteindelijke classificatie. Zoals ik ook in paragraaf 4.4.1 heb uitgelegd, wordt immers aan elk woord een gewicht toegekend – een zogeheten coëfficiëntwaarde. Met de kanttekening dat dit inzicht niet geheel representatief is voor de andere modellen, kunnen we, wanneer we de tien woorden met de hoogste en laagste coëfficiëntwaarde op een rij zetten (zie Figuur 4.7), enig inzicht in de woorden die volgens de computer karakteristiek – of juist totaal niet – zijn voor evaluatieve uitspraken. De figuur laat zien dat de woorden met de laagste coëfficiëntwaardes vooral uitblinken in nietzeggendheid: woorden als 'jou', 'eindelijk' of 'datgene' zou je intuïtief niet zozeer bij één van de twee categorieën plaatsen. Mogelijk worden niet-evaluatieve uitspraken eerder gekenmerkt door het *gebrek* aan evaluatieve woorden. De kenmerkende evaluatieve woorden zijn daarentegen een stuk makkelijk te interpreteren. We zien drie soorten woorden: (positieve) bijvoeglijk naamwoorden (boeiend, fraai, geloofwaardig), zelfstandige naamwoorden die aspecten van romans beschrijven (metafoor, stijl, humor, taalgebruik, dialogen) en twee enigszins curieuze auteursachternamen (die van J.M.H. Berckmans en Jaap Scholten – niet de meest canonieke auteurs uit de literatuurgeschiedenis). Het blijken dus, overeenkomstig met het model van Linders en Op de Beek, vooral woorden die specifieke aspecten (de zelfstandig naamwoorden) of eigenschappen (de bijvoeglijk naamwoorden) van romans aanduiden die volgens de computer kenmerkend zijn voor evaluatieve uitspraken.

Hoewel de computer evaluatieve uitspraken redelijk accuraat kan voorspellen, en de meeste modellen significant beter scoren dan een *random* baseline, zorgt de scheve verhouding tussen evaluatieve en niet-evaluatieve

uitspraken ervoor dat er nog percentueel nog relatief veel niet-evaluatieve uitspraken als evaluatief geclassificeerd worden. Tot op zekere hoogte liggen deze resultaten in lijn met de bevindingen van het onderzoek van Linders en Op de Beek. Ook tussen Linders en Op de Beek onderling was de overeenkomst over welke uitspraken nu daadwerkelijk evaluatief zijn aanvankelijk laag: bij hen werd in pilotonderzoek slechts 47.7% van de gevonden evaluaties door beide codeurs gemarkeerd.[166]

Ook wanneer we kijken naar de zinnen die de computer (op basis van het Nederlandse BERT-model) foutief als evaluatief heeft geclassificeerd, kunnen we allerminst zeggen dat elke zin zuiver descriptief van aarde is. Vergelijk:

(1) *Zwerm* is een roman die zich zowel in de vorm als in de inhoud verzet tegen het vastleggen van betekenissen, een roman die onbegrijpelijk wil zijn.

(2) Er komt een merkwaardig, verontrustend taalverloop (geen verhaalverloop, allerminst) tot stand, waarin weinig zeker is, alle mogelijke betekenissen verglijden.

(3) Het is niet te hopen dat zijn hier besproken roman een sleutelroman is, want dan staat het er maar beroerd voor in de wereld van de Nederlandse atoomgeleerden.

(4) Zowel Senior, Zoon, als Junior, hadden belangwekkende figuren kunnen worden, maar juist de autoritaire Zoon - een gestalte die Bordewijk toch voortreffelijk moet hebben' gelegen' - is bijzonder vaag gebleven.

Waar in (1) en (2) met 'onbegrijpelijk', 'merkwaardig', en 'verontrustend' nog bijvoeglijk naamwoorden gebruikt worden die nog niet direct de mening van de criticus weggeven – er zijn ook critici die juist een voorkeur hebben voor moeilijk, ongewoon taalgebruik – wijzen (3) en (4) met 'beroerd' en 'bijzonder vaag' al veel sterker in de richting van een evaluatieve uitspraak. Zuiver descriptief zijn deze zinnen in ieder geval niet. Of we het feit dat ze desalniettemin als niet-evaluatief zijn gemarkeerd kunnen verklaren doordat de codeurs, 'ondanks aandachtig lezen, sommige evaluaties over het hoofd zagen',[167] of dat het onderscheid tussen een descriptieve en een evaluatieve uitspraak inderdaad 'in geen enkel geval empirisch aantoonbaar' is,.[168] kunnen we in deze fase van het onderzoek niet meer achterhalen. Wel kunnen we vaststellen dat het classificeren van evaluatieve uitspraken voor mens én computer een bijzonder uitdagende taak is.

166 Hoewel ik er niet op uit ben de prestaties van het computermodel te cijfermatig af te zetten tegen die van menselijke codeurs, moet ik hier wel vermelden dat de maatstaf die Linders en Op de Beek hier hanteren strenger is dan *precision* en *recall*, omdat ze zowel *false negatives* als *false postives* meeneemt, maar niet, zoals bij *accuracy* dan het geval zou zijn, ook *true negatives*

167 Linders, "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005", 89.

168 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 110.

Tabel 4.5: Resultaten van de verschillende modellen en *features* voor het ranken van aspectcodes. Het model dat lichtrood gemarkeerd is, is het bestscorende model wanneer we kijken naar de *average precision*. BR = *Binary relevance.*

Model	Features	AP	One Error	Is Error	Ranking Loss
MF Baseline		0.51	0.65	0.77	0.25
BR-LR	BoW	0.70	0.40	0.55	0.13
	COWBIG	0.59	0.58	0.69	0.19
	LitR	0.74	0.35	0.51	0.11
	BOW + LitR	0.74	0.35	0.50	0.11
	BOW + COWBIG	0.71	0.39	0.54	0.13
Bi-GRU	LitR	0.75	0.33	0.48	0.11
	COWBIG	0.74	0.34	0.49	0.11
BERTMULTILINGUAL		0.77	0.32	0.45	0.10
BERTDUTCH		0.77	0.30	0.45	0.10

4.6.2 Ranken van aspectcodes

Tabel 4.5 toont de resultaten voor het ranken van aspectcodes, waarbij voor elk model de *Average Precision*, *One Error*, *Is Error* en de *Ranking Loss* (zie paragraaf 4.5.2) wordt weergeven. De verschillen tussen de modellen onderling zijn kleiner dan bij het classificeren van evaluatieve uitspraken. De meeste modellen zijn in staat de aspectcodes met een relatief hoge *average precision* – variërend van 0.70 tot 0.77 – te ranken. Het Nederlandse BERT-model scoort wederom op alle maatstaven het best, al is het verschil met het meertalige BERT-model beperkt. Over het algemeen verschillen de prestaties van de neurale netwerken niet uitzonderlijk veel dan die van de simpele regressiemodellen, hetgeen suggereert dat woordelijke sequenties hier minder informatief voor de relevante labels zijn dan in het binaire classificatieprobleem.

Hoewel ze wel overtroffen worden door de ontegenzeggelijke kracht van het Nederlandse BERT-model, leiden de beperktere domeinspecifieke *embeddings* consequent tot betere resultaten op ongeziene documenten dan de *embeddings* getraind op het gigantische COW-corpus. Het logistische regressiemodel op basis van COW-*embeddings* presteert op alle maatstaven zelfs veruit het slechts. Een interessant verschil ten opzichte van de vorige paragraaf is dat de LitR-*embeddings* hier wel tot betere classificaties leiden dan het regressiemodel op basis van BoW. De meest logische verklaring hiervoor is dat het aantal voorbeelden per label hier aanzienlijk beperkter is dan in het binaire classificatieprobleem (6.972 voorbeelden verdeeld over 12 labels tegenover 22.392 voorbeelden verdeeld over 2 labels). Hierdoor neemt de kans toe dat relevante maar mindergebruikte synoniemen niet in de trainingsset zitten, maar wel in de testset in de context van bepaalde labels voorkomen. En omdat het BoW-model geen rekening houdt met verhoudingen tussen woorden, kan het deze synoniemen niet extrapoleren naar de woorden die *wel* in de trainingsset zitten. Voor het BoW-model moeten we dus zo veel mogelijk taalvariatie al in onze voorbeelden hebben verdisconteerd, terwijl een dergelijke beperking voor (al dan niet contextuele) *word embeddings* die wél rekening houden met synoniemen, veel minder geldt. Voor datasets

Tabel 4.6: Resultaten van de verschillende modellen en *features* voor het ranken van eigenschappen. Het model dat lichtrood gemarkeerd is, is het bestscorende model wanneer we kijken naar de *average precision*. BR = *Binary relevance.*

Model	Features	AP	One Error	Is Error	Ranking Loss
MF baseline		0.49	0.69	0.75	0.23
BR-LR	BoW	0.63	0.49	0.64	0.15
	COW_{BIG}	0.54	0.62	0.73	0.20
	LitR	0.60	0.52	0.67	0.16
	BOW + LitR	0.63	0.47	0.63	0.15
	BOW + COW_{BIG}	0.63	0.49	0.64	0.15
Bi-GRU	LitR	0.67	0.42	0.58	0.14
	COW_{BIG}	0.68	0.41	0.57	0.12
$BERT_{MULTILINGUAL}$		0.67	0.41	0.58	0.14
$BERT_{DUTCH}$		0.69	0.38	0.55	0.12

met weinig voorbeelden per label zijn BoW-features dus aanzienlijk minder robuust.

4.6.3 Ranken van eigenschapscodes

De resultaten voor het ranken van eigenschapscodes – tenslotte – zijn te vinden in Tabel 4.6. De onderlinge verschillen resoneren tot op zekere hoogte met de verschillen tussen de modellen die we ook bij het ranken van aspectscodes zagen. Zo scoren de neurale modellen consequent beter dan de logistische regressiemodellen, en is het logistische regressiemodel op basis van COW-*embeddings* veruit het minst accuraat. Wederom scoort het Nederlandse BERT-model op alle maatstaven het best, al zijn de verschillen tussen de neurale modellen redelijk beperkt.

Hoewel de eigenschapscodes nog steeds redelijk accuraat kunnen worden geranked, blijkt over het algemeen dat het ranken van eigenschapscodes een aanzienlijk moeilijkere taak voor de computer is dan het ranken van aspectscodes: de *Average Precision* ligt hier ongeveer tien procentpunt lager dan in de vorige paragraaf. We kunnen hiervoor twee – niet exclusieve – oorzaken voor aanwijzen. De eerste is technisch: met vijftien verschillende eigenschapslabels is het aantal voorbeelden per label meer beperkt dan bij het ranken van aspectscodes, waar er keuze was uit twaalf verschillende labels. Ten tweede was de interbeoordelaarsbetrouwbaarheid bij Linders en Op de Beek voor het toekennen van aspectlabels voldoende tot goed (Cohen's $\kappa = .65$), terwijl deze voor de eigenschappen gematigd was (Cohen's $\kappa = .58$), wat er op duidt de eigenschapscodes over het algemeen minder eenvoudig eenduidig zijn toe te kennen dan de aspectcodes.

Hoofdstuk 5

CORPUS

Voor de vergelijking van Nederlandse en Vlaamse literatuurkritiek heb ik een substantiële verzameling van boekrecensies – gepubliceerd tussen 1 januari 1946 en 31 december 2018 – aangelegd. De recensies heb ik geraadpleegd via *LiteRom*.[169] Hoewel NBD Biblion – de organisatie achter het digitale archief – zelf aangeeft dat *LiteRom* een '*compleet* [mijn cursivering, JV] overzicht van auteursinterviews, artikelen en boekrecensies van Nederlandse [...] literatuur' biedt, heeft onder andere Esther op de Beek steekproefsgewijs beargumenteerd dat 'lang niet alle recensies' in de database te vinden zijn.[170] Toch beschouw ik *LiteRom* als een geschikte bron voor de verzameling van mijn corpus. *LiteRom* mag dan niet helemaal compleet zijn, het is wél, in de formulering van Jos Joosten, 'the *most* complete Dutch database of book reviews'.[171] Streven naar compleetheid is mijns inziens altijd aan te moedigen, maar statistisch allerminst noodzakelijk. Het gaat mij in deze scriptie vooral om het verzamelen van een corpus dat substantieel genoeg is om generaliserende uitspraken te kunnen doen over Nederlandse en Vlaamse literatuurkritiek, en het grote aantal recensies op *LiteRom* lijkt mij daar absoluut voldoende geschikt voor.

LiteRom bevat recensies van zowel oorspronkelijk Nederlandstalige als vertaald proza, poëzie en non-fictie, uit landelijk en regionale dag-, week- en maandbladen, overwegend gepubliceerd in Nederlandstalige bronnen, maar ook met een beperkt aantal buitenlandse recensies.[172] Zoals ik in de inleiding al heb aangegeven, beperk ik mij in deze scriptie alleen tot recensies die verschenen zijn in Nederlandstalige landelijke dagbladen en oorspronkelijk Nederlandstalig fictioneel proza bespreken. Het classificatiemodel kan namelijk niet eenvoudigweg gegeneraliseerd worden naar recensies van een andere aard – voor poëzierecensies zouden bijvoorbeeld al heel andere aspecten en eigenschappen (zoals: ritme en klank) van belang kunnen zijn.

169 https://literom.nbdbiblion.nl

170 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 81.

171 Jos Joosten, "6 A Big Bang? The Paradox of the Reception of Günter Grass's Die Blechtrommel in The Netherlands", in *The Echo of Die Blechtrommel in Europe* (BRILL, 2016), 121.

172 Bijvoorbeeld een recensie van Mulisch' *Discovery of Heaven* in *The Sunday Times*: https://literom-nbdbiblion.nl/detail/321196/the-blue-eyed-boy

5.1 DATAVERZAMELING

De interface van *LiteRom* biedt slechts beperkte mogelijkheden om het corpus op de voor ons relevante criteria te selecteren. We kunnen de zoekresultaten weliswaar beperken aan de hand van de publicatiedatum van de recensie en de oorspronkelijke taal van het besproken boek, maar verdere informatie over de afkomst van de bron of het genre van het boek is niet – ter selectie of anderszins – beschikbaar. Evenmin kunnen we in *LiteRom* de herkomst van de recensent achterhalen, hetgeen natuurlijk een van de belangrijkste variabelen van dit onderzoek is. Ik heb daarom het corpus verder gefilterd en verrijkt aan de hand een combinatie van automatische en handmatige methodes. Daarvoor ben ik begonnen met de selectie van recensies die ik wél in *LiteRom* kon maken, namelijk de recensies over oorspronkelijk Nederlandstalige boeken gepubliceerd tussen 1946 en 2018. Dit resulteerde in een totaal van 72.077 recensies.[173] In wat volgt ga ik stapsgewijs in op de verdere selectie- en verrijkingsmethodes.

5.1.1 Kranten

Ten eerste heb ik het aanvankelijke corpus van 72.077 recensies gefilterd op basis van de bron van de recensie. Hiervoor heb ik alle 351 verschillende brontitels handmatig, via de eerdergenoemde *Doccano*-tool, gecodeerd op twee aspecten: het type publicatie (nationaal dagblad, periodiek, of regionaal dagblad) en de herkomst van de publicatie (Nederlands, Vlaams, overig). Ik selecteer alle recensies die gepubliceerd zijn in een Nederlands of Vlaams nationaal dagblad – een verdere selectie van titels maak ik niet. Wel heb ik, deels omdat Vlaamse recensies in *LiteRom* schaarser zijn, voor Vlaamse recensies het begrip 'regionaal' iets ruimer genomen, en *Belang Van Limburg* en *Gazet Van Antwerpen*. Deze kranten richten zich weliswaar op een specifieke regio, maar hebben in Vlaanderen een marktaandeel vergelijkbaar met dat van de *Volkskrant* nu in Nederland.[174]

Een keuze voor titels die over de hele periode hebben gepubliceerd, zoals Linders en Op de Beek die maken, is misschien de meest logische optie wanneer we ontwikkelingen in die specifieke titels willen traceren, maar wat mij betreft veel minder wanneer we over 'de' literatuurkritiek willen spreken. In dat geval werkt een selectie van titels die nu nog publiceren juist bijzonder anachronistisch, omdat ze impliceert dat de titels die nu relevant zijn in het verleden ook een dergelijke stempel op de literatuurkritiek hebben gedrukt. Zo worden de recensies die Kees Fens in het katholieke dagblad *De Tijd* schreef door Linders en Op de Beek niet meegenomen, alleen omdat

173 En dit ook de selectie aan recensies die ik heb gebruikt om mijn *word embeddings*-model (LiTR) te trainen. Zie paragraaf 4.3.2

174 Zie Johan Van Croonenborch, "De prijs van het publieke debat: een bedrijfseconomische geschiedenis van de Vlaamse krantenpers, 1830-1980" (proefschrift, Ghent University, 2018), voor de historische ontwikkelingen in het marktaandeel van Vlaamse dagbladen; zie Vlaamse Regulator voor de Media.(VRM), *Mediaconcentratie in Vlaanderen. Rapport 2019*, 2019, voor meer recentere data.

het dagblad later is opgeheven.[175] Een overzicht van alle titels – en de jaren waarin recensies van die titels in *LiteRom* staan – is te vinden in Figuur 5.1.

Van alle 72.077 recensies in *LiteRom* blijken er 14.315 gepubliceerd in periodieken, 18.321 in regionale dagbladen, 35.920 in nationale dagbladen, en van 3.521 recensies was de herkomst (van de bron) onbekend. In de laatste categorie vallen zowel recensies waarvan er geen bron vermeld stond in *LiteRom* als recensies waarvan de brontitel niet eenvoudig te herleiden was naar een specifieke publicaties, titels zoals *Oogst*, *Bode van nu* of *Kritische Aantekeningen*.[176] Van de 35.920 recensies gepubliceerd in nationale dagbladen zijn er 8.948 gepubliceerd in Vlaamse dagbladen, 26.923 in Nederlandse dagbladen en slechts 49 in een niet-Nederlandstalig dagblad. Dit resulteert in een totaal van 35.871 relevante recensies.

5.1.2 Genres

In deze 35.871 recensies worden maar liefst 15.904 verschillende boeken besproken. Dit zijn niet alleen prozawerken, maar ook poëziebundels en nonfictieboeken, en er is in *LiteRom*, zoals gezegd, geen enkele manier om het genre (in de breedste zin) van het besproken boek te achterhalen – we weten alleen de titel en de auteur van het boek, en niet de ISBN of iets soortgelijks. Omdat het aantal besproken werken óók veel te groot is om handmatig te coderen, heb ik ervoor gekozen dit proces te automatiseren. Twee stappen bleken hiervoor noodzakelijk. Ten eerste heb ik het besproken boek proberen te koppelen aan een vermelding in de database van de (Nederlandse) Koninklijke Bibliotheek. Hiervoor heb ik gebruik gemaakt van publieke API van de KB.[177] Omdat de zoekfunctie van de KB-API vrij streng is,[178] leverde dit doorgaans maar één resultaat op. Wanneer er toch meerdere resultaten gevonden werden (als een boek meerdere drukken heeft, of bijvoorbeeld ook een stripboekbewerking), heb ik alle resultaten geselecteerd waarvan de boektitel en de achternaam van de auteur in de KB-database het minst verschillen met die in *LiteRom*.[179] Vervolgens heb ik voor alle resultaten de collectiebeschrijving (Trefwoord Depot, Trefwoord GOO, Basisclassificatie) doorzocht op het voorkomen van de trefwoorden 'roman', 'proza', '301' 'novelle' of 'verhalenbundel', om het genre van het besproken werk te extraheren.

Van 13.021 van de 15.904 besproken werken kon op deze wijze een vermelding in de KB-database gevonden worden. De recensies van de geï-

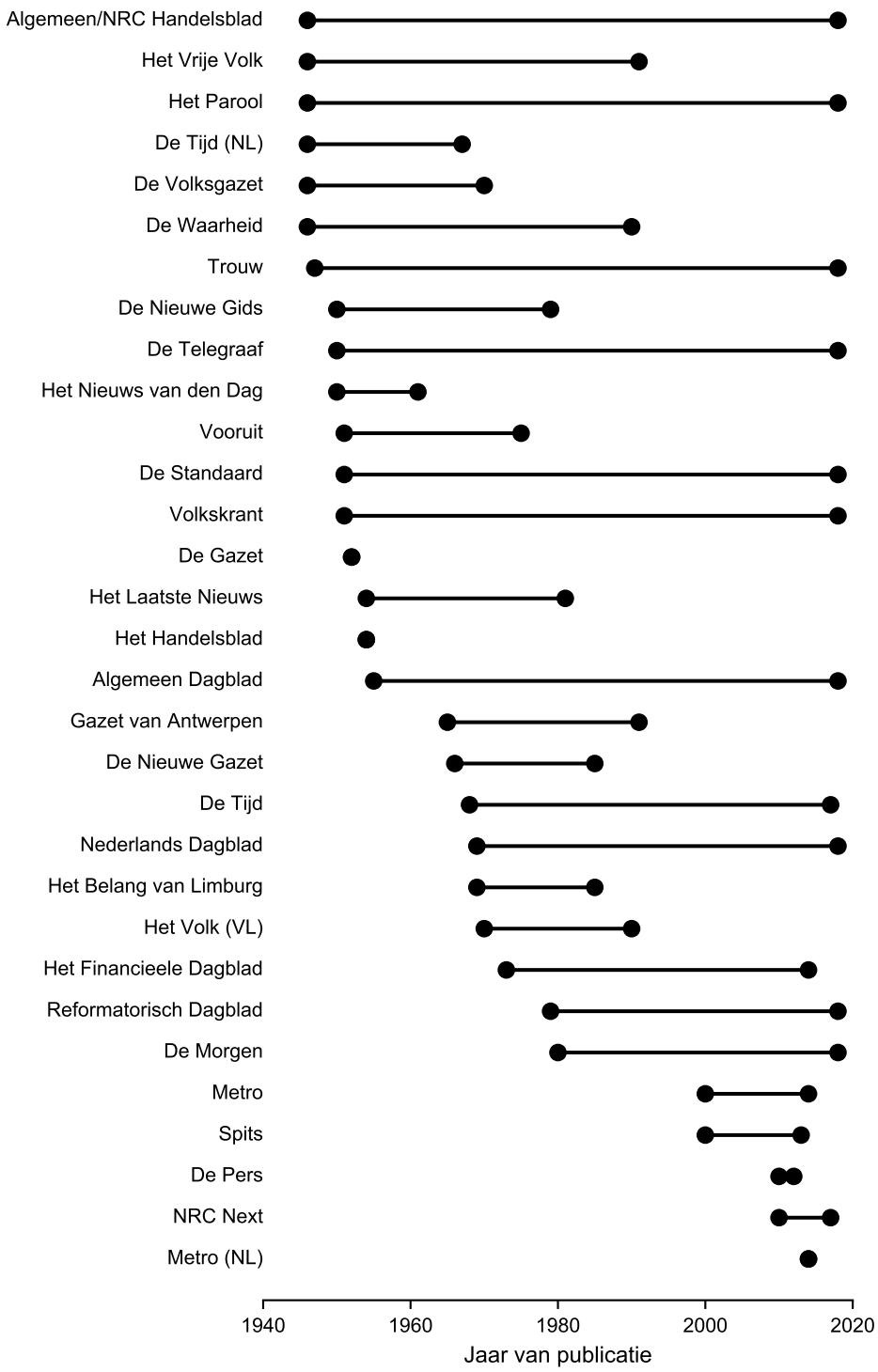
175 Dit wordt extra problematisch wanneer Kees Fens voor een groot deel verantwoordelijk is voor een toename in structuurevaluaties. Vgl. Vitse, "Esther op de Beek, Een literair fenomeen van de eerste orde. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: Een kwantitatieve en kwalitatieve analyse.", 151.

176 Dergelijke titels zijn uiteraard waarschijnlijk wel te herleiden met een uitgebreide archiefstudie, maar vanwege hun beperkte aantal (minder dan de vijf procent van het totaal) en de nog veel beperktere kans dat deze obscure titels *nationale* dagbladen waren, loont dat voor dit onderzoek de moeite niet.

177 Zie `http://jsru.kb.nl/sru/sru?operation=explain`

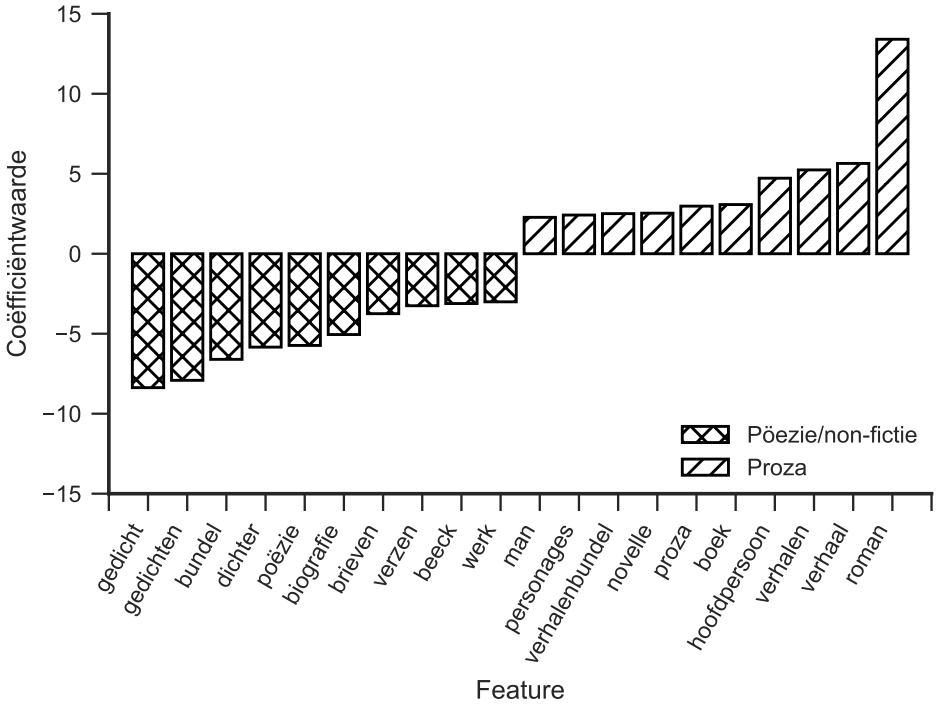
178 Zo levert een combinatie van de titel *De Avonden* en de foutief getypte auteursnaam 'Teve' geen resultaten op. Zie `http://jsru.kb.nl/sru?x-collection=GGC&version=1.1&operation=searchRetrieve&query=%22De%20Avonden%22%20and%20dc.creator=%22Teve%22`

179 Hiervoor heb ik, heel simpel, de Levenshteinafstand van de auteursnaam en de boektitel in de KB-database en *LiteRom* gedeeld door twee.



Figuur 5.1: De temporele reikwijdte van alle kranten in het recensiecorpus.

dentificeerde werken zijn vervolgens in de tweede stap gebruikt om een classificatiemodel te trainen voor het classificeren van de recensies van de werken waarvan het genre nog niet bekend was. Hiervoor heb ik een simpel logistisch regressiemodel op basis van *Bag-of-Words*-features getraind. Zonder al te veel hyperparameteroptimalisatie behaalde dit model een F_1-score



Figuur 5.2: De top 10 woorden met de hoogste (kenmerkend prozarecensies) en laagste (kenmerkend voor poëzierecensies) coëfficiëntwaarde in het logistische regressiemodel op basis van de bag-of-words-representaties. Deze woorden stuurden het model het sterkst richting één van de twee labels.

van 0.85. Omdat het criterium van correctheid in mijn onderzoek zwaarder weegt dan het criterium van compleetheid, heb ik de *probability cut-off* verhoogt naar 0.64, hetgeen resulteert in een *precision* van 0.94 (en een *recall* van 0.81) voor het classificeren van prozarecensies. Voor de werken waarvan er meerdere recensies waren heb ik de klasse bepaald aan hand van de gemiddelde waarschijnlijkheid over alle recensies van dat werk. Van de 35.871 zijn er zodoende 22.705 geclassificeerd als prozarecensies, en 13.166 als non-fictie/poëzierecensies. Figuur 5.2 geeft inzicht in de woorden die volgens de computer karakteristiek bleken voor prozarecensies.

5.1.3 Herkomst van recensent

Hoewel we reeds de herkomst van de kranten hebben geïdentificeerd, kies ik er in deze scriptie voor om het adjectief in 'Vlaamse' of 'Nederlandse' literatuurkritiek primair te operationaliseren aan de hand van de nationaliteit van recensent, en niet aan de hand van de herkomst van de bron. Dit omdat ik het literaire en culturele milieu waarin individuele critici zich vormen representatiever acht voor nationale verschillen dan de institutionele context van de recensie. Dat betekent overigens niet dat kranten hier in mijn optiek helemaal geen betekenisvolle rol in spelen: ook kranten kunnen een – bij gebrek aan een betere term – *hub* voor recensenten met een gelijkaardige culturele of ideologische achtergrond vormen, en in die zin zou een hypothetische Ne-

derlands criticus die in de jaren zestig de evenzeer hypothetische ontzuiling van de Nederlandse literatuurkritiek ontvlucht en voor een Vlaamse krant gaat schrijven, wellicht eerder representatief zijn voor de aard Vlaamse literatuurkritiek. Toch verwacht ik dat dergelijke gevallen uitzonderlijk zijn, en dat het culturele milieu – al dan niet op de universiteit gevormd – alsnog zwaarder meeweegt.

Daarom heb ik voor alle 2.241 recensenten de herkomst, wederom met behulp van de *Doccano*-tool, handmatig gecodeerd. Voor het achterhalen van de nationaliteit van de recensent heb ik, wanneer deze voor mij niet gelijk evident was, in de eerste plek publieke bronnen – *DBNL*, *Het Biografisch Portaal van Nederland*, *Wikipedia* – geraadpleegd. De nationaliteit van het overgrote deel van de recensenten kon hiermee achterhaald worden. Voor het overige deel, voor de recensenten waarvan de biografische informatie niet direct beschikbaar was, waaronder de recensenten die alleen onder hun initialen in *LiteRom* vermeld stonden, heb ik de nationaliteit alsnog herleid aan de hand van de herkomst van de krant.[180]).

Van de 22.705 proza-recensies zijn er 17.799 geschreven door Nederlanders, 4.895 door Vlamingen en de resterende 11 door critici met een andere nationaliteit. Na deze laatste stap in het selectieproces blijven er van de 72.077 aanvankelijke recensies nog 22.694 recensies over die aan alle criteria voldoen. Figuur 5.3 geeft een overzicht van het volledige selectie- en filterproces van het corpus.

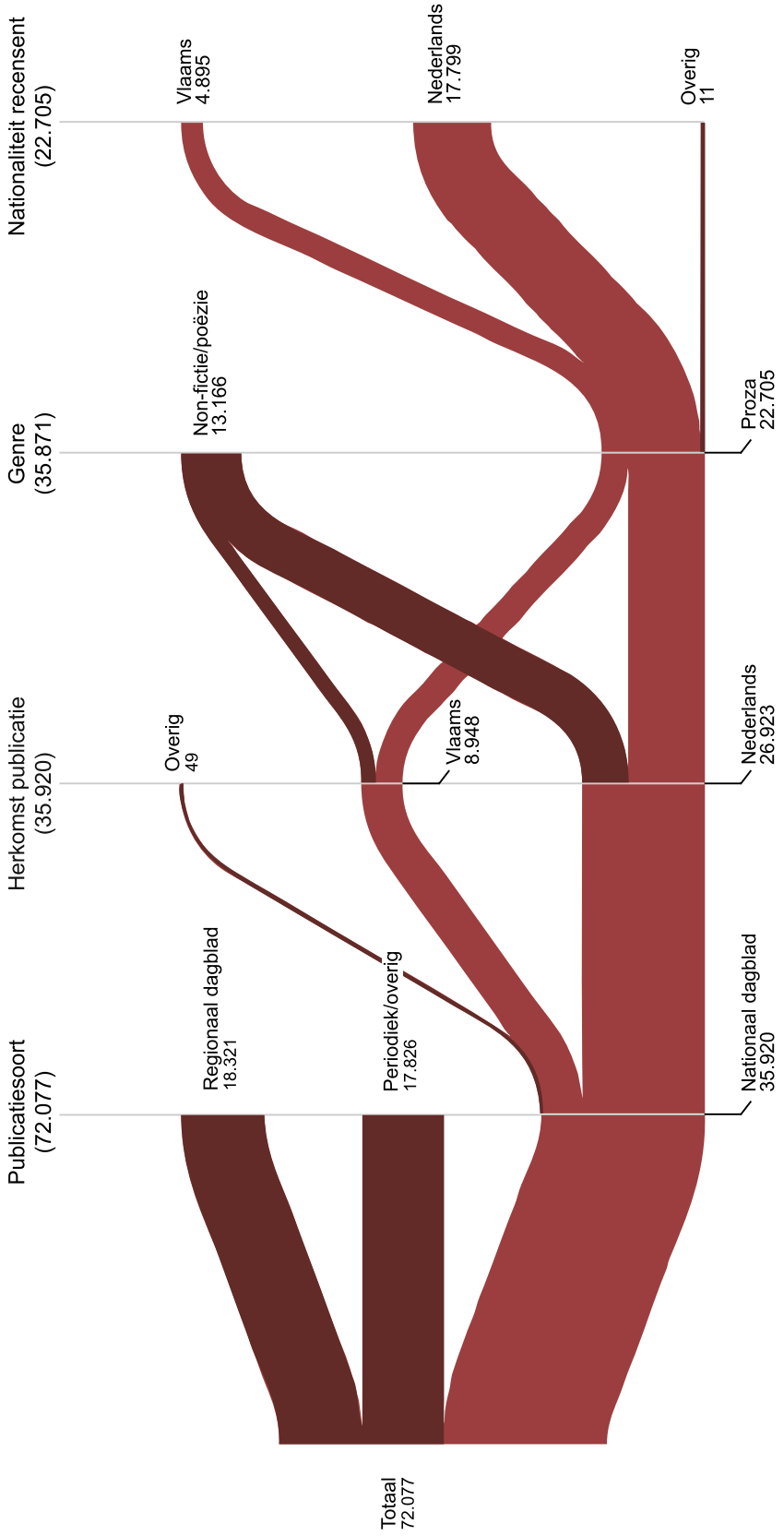
Als we de herkomst van de krant in een kruistabel afzetten tegen de herkomst van de recensent (zie Tabel 5.1), zien we dat Vlaamse en Nederlandse critici – niet geheel verrassend – overwegend binnen hun eigen landgrenzen schrijven. Het overgrote merendeel van recensies geschreven door Nederlandse critici is gepubliceerd in een Nederlandse krant, en het meerendeel van de recensies van Vlaamse critici in Vlaamse kranten. Een chikwadraattoets laat zien dat Nederlandse en Vlaamse critici evenmin verschillen in de mate waarin ze voor een zogezegd 'buitenlandse' krant schrijven: recensies in Vlaamse dagbladen zijn niet significant vaker door een Nederlandse recensent geschreven dan Nederlandse recensies door een Vlaamse criticus ($\chi^2 = 0.87, p = 0.35$).[181]

5.1.4 Herkomst van de besproken schrijver

Ten slotte heb ik ook de herkomst van de schrijvers van de besproken boeken gecodeerd. Omdat de selectie voor oorspronkelijk Nederlandstalig werk wél al in *LiteRom* kon worden gemaakt, is dit formeel geen onderdeel van het selectieproces. Het coderen van deze variabele dient vooral een analytisch

180 Van de 22.694 recensies waren er 1.162 (5.12%) recensies waarin de auteur in *LiteRom* ofwel onbekend was, ofwel alleen onder zijn of haar initialen vermeld stond (recensies van 'K.F.' in *De Tijd* en de *Volkskrant* uitgezonderd). In het uiterste, bijzonder onwaarschijnlijke, geval dat de herkomst van deze recensenten tegenovergesteld is aan de herkomst van de krant waarvoor ze schrijven, levert deze keuze dus een relatief beperkte ruis in de dataset op.

181 Voor deze berekening heb ik niet bovenstaande kruistabel gebruikt, maar heb ik een derde variabele aangemaakt, waarbij de variabele de waarde 'Land van herkomst' aanneemt wanneer het land van de criticus overeenkomt met het land van de krant, en de waarde 'Buitenlands' wanneer dit niet het geval is.



Figuur 5.3: Een visueel overzicht van het volledige selectieproces van het corpus. De rode lijn die uitmonden bij Vlaams en Nederlands representeren het uiteindelijke corpus. In totaal telt het corpus van dit onderzoek dus 22.694 recensies.

Tabel 5.1: Kruistabel van de herkomst van recensent en de herkomst van de krant.

Herkomst recensent	Land van krant		
	Nederlands	Vlaams	Totaal
Nederlands	17253	546	17799
Vlaams	163	4732	4895
Totaal	17416	5278	22694

doel: zodoende kunnen we vergelijken of Nederlandse schrijvers op een andere manier beoordeeld worden dan Vlaamse schrijvers.

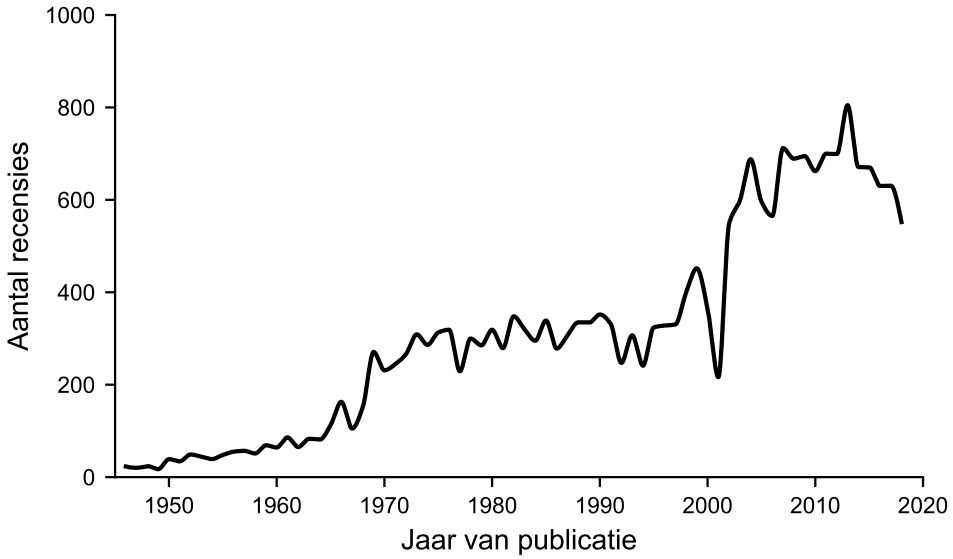
Ik heb alle 3010 unieke schrijvers in ons recensiecorpus gecodeerd via de *Docanno*-interface. Tijdens het coderen bleek de selectiemethode – zowel die van *LiteRom* als die van mij – niet volledig waterdicht. Zo zaten er zowel buitenlandse auteurs als auteurs die alleen *thrillers* of *chick lit* schrijven tussen.[182] Dergelijke gevallen heb ik gecodeerd als 'Nederlands' noch 'Vlaams', maar als 'overig'. Dit bleek het geval voor 116 van de 3010 auteurs, die samen goed zijn voor 333 recensies. Deze heb ik vervolgens alsnog uit het corpus gefilterd. Daarmee telt het recensiecorpus in totaal 22.361 recensies, waarvan er in 4.351 een Vlaamse roman besproken wordt, en in 18.010 een Nederlandse.

5.2 CORPUSSTATISTIEKEN

Na het toepassen van alle selectiecriteria op de teksten die in *LiteRom* te vinden zijn, zoals beschreven in de vorige paragraaf, telt het corpus van Nederlandse en Vlaamse dagbladrecensies dat ik in dit onderzoek gebruik 22.361 recensieteksten. Zoals Figuur 5.4 laat zien, is dat aantal niet evenwichtig per jaar verdeeld. Voor de jaren tot 1950 bevat het corpus maar ongeveer 20 teksten per jaar. Dit aantal loopt vanaf 1950 tot 1970 gestaag op van 40 naar ongeveer 150 teksten per jaar, overeenkomstig met de omvang van de jaren van het handmatig gecodeerde recensiecorpus van Linders en Op de Beek. Na 1970 neemt het aantal recensieteksten flink toe, met gemiddeld 350 recensies per jaar, en ook na 2000 kunnen we een relatief snelle toename van het aantal recensies zien, gemiddeld ongeveer 700 per jaar, met maar liefst 805 recensies van Nederlandstalig proza in 2013 – een groter aantal dan de *totale* omvang van het corpus van Linders en Op de Beek.

Hoewel het mogelijk enigszins voor de hand ligt, acht ik het belangrijk om te benadrukken dat deze explosieve, haast exponentiële, stijging van het aantal recensies per jaar geenszins representatief voor de daadwerkelijke productie van recensies. Linders en Op de Beek, die voor de samenstelling van hun corpus daadwerkelijk de kranten in zijn gedoken om een accurate indruk te krijgen van het aantal recensies, rapporteren weliswaar een stijging

182 Slechts 174 van de 22.705 recensies bespreken een boek dat geschreven is door auteurs in deze tweede categorie, waarvan er maar liefst 148 gepubliceerd zijn na 2000. Omdat deze genres maar een zeer beperkt deel uitmaken van het totaal, en ik verwacht dat, net als poëzierecensies, dat voor *thrillers* en *chick lit* wezenlijk andere criteria van belang zijn (denk alleen al plot), heb ik deze genres uitgesloten van mijn analyse.



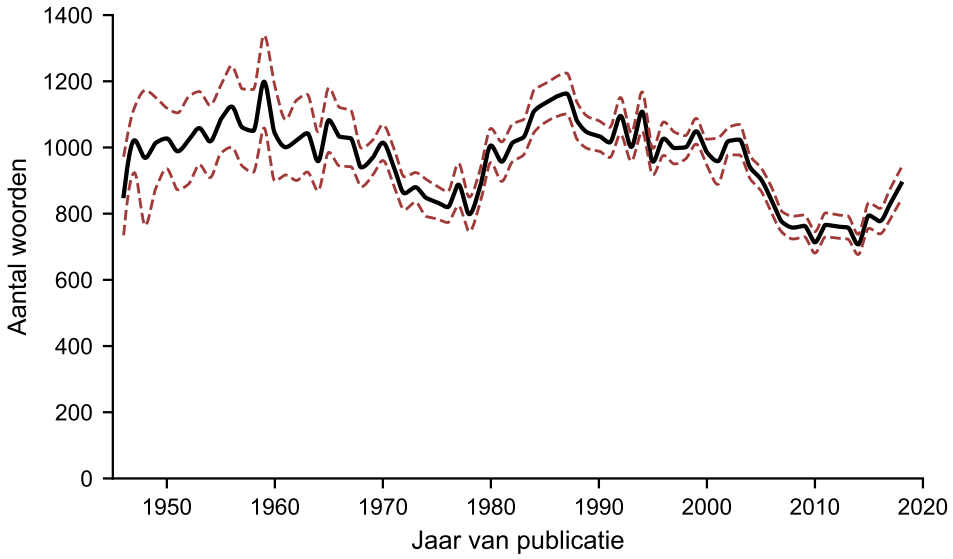
Figuur 5.4: Het totale aantal recensies per jaar in mijn uiteindelijke recensiecorpus.

in het aantal recensies van 1955 tot 2005, maar dat is hoogstens een verviervoudiging – geenszins de verachtienvoudiging die we in Figuur 5.4 aantreffen. Uiteindelijk zijn de aantallen die ik hier rapporteer dan ook vooral representatief voor het aantal recensies per jaar die in *LiteRom* te vinden zijn.

Dat het corpus onevenwichtig is verdeeld, is echter niet noodzakelijk problematisch voor de diachrone analyse. Uiteraard moeten de frequenties van aspect- en eigenschapscodes gecorrigeerd worden voor het aantal recensies per jaar, en zijn de cijfers van de jaren met meer recensies betrouwbaarder dan die van de jaren met een beperkt aantal recensies, maar over het algemeen is de steekproefgrootte per jaar meer dan voldoende om er generaliserende uitspraken op te baseren. Dat geldt in minder mate voor de eerste twintig jaar van het corpus, en dan in het bijzonder voor het Vlaamse aandeel in het aantal recensies, met enkele jaren waarin het aantal recensies van Vlaamse hand minder dan vijf telt. Daarom gebruik ik voor de diachrone analyse van deze scriptie *Gaussian Process smoothing* – een techniek om nieuwe waardes te benaderen voor de geobserveerde datapunten zodat deze zich meer effen over de x-as verplaatsen – om de gemiddeldes per jaar te corrigeren voor individuele uitschieters, en meer algemene trends en ontwikkelingen te traceren.[183]

Naast het aantal recensies fluctueert ook het gemiddeld aantal woorden per jaar (zie Figuur 5.5). Zoals we kunnen zien in de figuur, lijkt het gemiddeld aantal woorden tussen 1946 en 1960 lichtelijk toe te nemen. Daarna neemt dit aantal flink af: waar in 1960 de recensies gemiddeld iets meer dan 1000 woorden tellen, zijn deze rond 1975 meer dan 200 woorden korter. Na 1975 is er echter weer een substantiële stijging waarneembaar in ons recensiescorpus, waarbij het aantal woorden redelijk stabiel blijft tot het jaar 2000. Daarna neemt het gemiddeld aantal woorden wederom behoorlijk af, met

183 Voor een uitgebreide uitleg van de werking van Gaussian Process smoothing, zie Audrey Kuzmenko, *Gaussian Process (GP) smoothing*, `https://docs.pymc.io/notebooks/GP-smoothing.html`, 2018.



Figuur 5.5: De ontwikkeling van het gemiddelde aantal woorden per jaar in recensies. De rode lijnen geven de 95% betrouwbaarheidsinterval aan.

het dieptepunt in 2014, waarin de recensies gemiddeld 707 woorden telden. Na 2014 lijkt er wederom een lichte stijging waarneembaar. Zo tellen de recensies in 2018 – het meest recente jaar in het corpus – gemiddeld alweer 891 woorden. Tegelijkertijd is deze stijging in de afgelopen vier jaar te beperkt om te kunnen zeggen of ze indicatief is voor een langere trend.

Hoofdstuk 6

ANALYSEMETHODE

Met het classificatiesysteem voldoende accuraat en het recensiecorpus compleet beschrijf ik in dit hoofdstuk de laatste methodologische stappen die ik heb ondernomen vooraleer ik mijn daadwerkelijke analyse kan aanvangen. Ik beschrijf op welk analytisch niveau ik opereer, de manier waarop ik de gemiddeldes bereken en vergelijk, de statistische toetsen die ik daarop uitvoer, en de wijze waarop de resultaten van deze toetsen interpreteer.

Allereerst heb ik het drieledige recensieclassificatiesysteem toegepast op de 22.361 teksten in mijn recensiecorpus, waarbij ik voor alle drie classificatietaken gebruik heb gemaakt van het Nederlandstalige BERT-model, dat immers voor alle taken het meest accuraat bleek. Daarvoor heb ik alle recensies in zinnen gesplitst met *Ucto*,[184] resulterend in een totaal van 1.107.486 zinnen. Deze zinnen heb ik vervolgens gevoerd aan het eerste BERT-model om de evaluatieve uitspraken te classificeren. 290.782 zinnen zijn zodoende als evaluatief geclassificeerd, en deze set heb ik aan het tweede en derde BERT-model gevoerd om de aspect- en eigenschapscodes te voorspellen.

6.1 ANALYSENIVEAU

Gegeven een zin s die door het eerste model als evaluatief is geclassificeerd, retourneren de tweede en derde BERT-modellen voor elk label λ in de labelset $\mathcal{L}$ (alle aspect- of eigenschapscodes) de kans dat de gegeven zin met label λ wordt geassocieerd, ofwel $\hat{p}(\lambda|s)$. Deze kansen, lopend van 0 tot 1, vormen de elementaire bouwstenen van mijn analyse – hierop baseer ik al mijn verdere berekeningen. Hoewel ik er ook voor zou kunnen kiezen om de labels daadwerkelijk aan de zinnen toe te kennen wanneer de kans hoger is dan een bepaalde kritieke waarde (conventioneel: $\hat{p} > 0.5$), of mijn analyse alleen te baseren op de ordening van de labels, biedt deze meer probabilistische benadering mijns inziens het best de mogelijkheid de fluïditeit van de aspect- en eigenschapscategorieën in mijn analysemodel te exploiteren (zie ook paragraaf 4.2). De categorieën worden zodoende immers niet zozeer op binaire wijze wel/niet met de zinnen geassocieerd, maar kennen *gradaties* van associatie.

184 https://languagemachines.github.io/ucto/

Vervolgens bereken ik op recensieniveau voor elk label het gemiddelde aandeel. Meer formeel definieer ik $\hat{p}(\lambda|d)$ als de kans dat een willekeurige recensie d met label λ geassocieerd wordt. Ik benader $\hat{p}(\lambda|d)$ aan de hand van het genormaliseerde gemiddelde over alle evaluatieve uitspraken in recensie d:

$$\hat{p}(\lambda|d) = \frac{\sum_{s \in S_1^d} \hat{p}(\lambda|s)}{\sum_{\lambda_i \in \mathcal{L}} \sum_{s \in S_1^d} \hat{p}(\lambda_i|s)}. \tag{10}$$

Waarbij S_1^d de set van alle evaluatieve uitspraken in document d is, en $\sum_{\lambda_i \in \mathcal{L}}$ gebruikt wordt om de waardes te normaliseren over alle labels in de labelset. Ik normaliseer de eigenschapscodes apart van de aspectcodes. Zodoende tellen voor elke zin de waardes van de aspectcodes op tot 1, en tellen de waardes van eigenschapscodes eveneens op tot 1. Een alternatieve strategie zou zijn om de teller in vergelijking (10) te delen door het aantal evaluatieve uitspraken, ofwel $|S_1^d|$. Deze manier van berekenen zou echter niet robuust voor verschillen in zinslengte zijn: recensies met langere zinnen zouden zodoende – ervan uitgaande het aantal aspect- en eigenschapscodes toeneemt met de lengte van een zin – aanmerkelijk hogere waardes hebben dan recensies met kortere zinnen. De uitkomst van (10) gebruik ik vervolgens weer om de gemiddeldes voor de Vlaamse en Nederlandse literatuurkritiek als geheel, geoperationaliseerd aan de hand van de herkomst van de recensent, en de gemiddeldes per jaar te berekenen.

Hoewel we met deze meer probabilistische benadering in mijn optiek beter in staat zijn de fluiditeit van de aspect- en eigenschapscodes in de analyse te verwerken, is het analyseren van gemiddeldes op deze wijze – of het analyseren van gemiddeldes *an sich* – ook niet geheel zonder problemen. Door de labels op recensieniveau te analyseren, verliezen we informatie over de mate waarin labels op zinsniveau voorkomen. Zo kan een label met een relatief beperkte waarschijnlijkheid over de gehele recensie hetzelfde gemiddelde krijgen als een label dat in enkele zinnen heel prominent is, maar in de rest van de recensie helemaal niet. Ik kies toch voor het berekenen van gemiddeldes, en niet voor de analyse een zinsniveau, om twee redenen. Ten eerste is het niet zozeer mijn doel om de prominentie van labels *onderling* te vergelijken, maar wil ik vooral weten of er verschillen zijn *tussen* Vlaamse en Nederlandse recensies in het gewogen aandeel van deze labels. Ten tweede, en nog veel belangrijker, probeer ik de analysemethode van mijn onderzoek zo veel mogelijk vergelijkbaar te houden met het onderzoek van Linders en Op de Beek, die immers ook met gemiddeldes per recensie werken.[185] De gemiddelde waarschijnlijkheid van een label is weliswaar iets anders dan het procentuele aandeel hiervan, maar als we naar de distributies van deze waardes over de recensies kijken (zie Bijlage A.3 voor een compleet overzicht), dan lijkt mijn probabilistische benadering over het algemeen een

185 De mogelijkheden dit onderzoek met dat van Linders en Op de Beek goed te kunnen vergelijken biedt twee evidente voordelen. Ten eerste kan ik verder valideren in hoeverre mijn model, buiten een experimentele context, accurate voorspellingen maakt. Als de resultaten voor het Nederlandse deel al te radicaal afwijken van die van Linders en Op de Beek, biedt dat reden tot scepsis ten opzichte van de validiteit van mijn model. Als, andersom, mijn resultaten niet zozeer radicaal afwijken, maar eerder op subtiele doch betekenisvolle punten, kunnen we de resultaten van Linders en Op de Beek kritisch bevragen.

(beperkt) democratisch effect te hebben op de onderlinge verhoudingen tussen labels: minder frequente labels, die de modellen over het algemeen ook voorzichtiger classificeren, hebben een evenwichtiger gemiddeld aandeel in de recensies. Ook de invloed van hoogfrequente labels neemt lichtelijk af. Tegelijkertijd zien we dat de gemiddelde waarschijnlijkheid per recensie van de veelvoorkomende labels in mijn benadering nog slechts zelden tegen de nul aangrenst, hetgeen inderdaad een structureel – maar niet altijd betekenisvol – aandeel in de recensies lijkt aan te duiden.

6.2 STATISTISCHE TOETSEN

Omdat de data niet voldoen aan de assumpties voor parametrische toetsen,[186] toets ik de verschillen tussen Nederlandse en Vlaamse literatuurkritiek non-parametrisch. Meer specifiek voer ik, wederom evenals Linders en Op de Beek, Mann-Whitney toetsen uit om twee groepen onderling te vergelijken. Wanneer ik meer dan twee groepen vergelijk – bijvoorbeeld wanneer ik naast de herkomst van de recensent ook die van de schrijver in mijn analyse verdisconteer – voer ik eerst een Kruskal-Wallistoets om te bepalen of er een hoofdeffect tussen groepen is. Wanneer dat het geval is, voer ik Mann-Whitney toetsen uit om alle groepen paarsgewijs te vergelijken, waarbij ik door middel van een Bonferroni-correctie corrigeer voor herhaalde toetsen. De toetsen voer ik uit op de gemiddeldes per *recensie*, ofwel $\hat{p}(\lambda|d)$ (zoals berekend in de vorige paragraaf). Daarbij gebruik ik het Python-pakket *SciPy* om de toetsen uit te voeren.[187]

6.3 EFFECTGROOTTES

Naast de teststatistieken en de p-waardes van de Mann-Whitneytoetsen rapporteer ik ook de bijbehorende effectgroottes. Dat een verschil significant is, betekent in de context van deze scriptie namelijk nog niet noodzakelijk dat het verschil ook bijzonder betekenisvol is: met meer dan 20.000 observaties kunnen ook vrij kleine verschillen significant zijn. Waar de p-waardes alleen vertellen hoe groot de kans is dat een gelijkaardig verschil voorkomt wanneer er *geen* verschil tussen groepen zou zijn, bieden effectgroottes ons de mogelijkheid de *grootte* van dat verschil op gestandaardiseerde wijze te duiden. Ik gebruik hiervoor, zoals voorgesteld door Andy Field, Pearson's r als schatting van de effectgrootte, hetgeen in de context van Mann-Whitneytoetsen berekend kan worden door de z-score van de teststatistiek te delen door het aantal observaties:[188]

$$r = \frac{z}{\sqrt{N}}. \tag{11}$$

186 Linders, "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005", 114.

187 Pauli Virtanen en andere, "SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python", *arXiv e-prints*, arXiv:1907.10121, juli 2019, arXiv:1907.10121, arXiv: 1907.10121 [cs.MS].

188 Andy Field en andere, *Discovering statistics using R* (Sage publications, 2012), 665.

Pearson's *r* loopt van 0 tot 1, waarbij 0 totaal geen effect representeert, en 1 een perfect effect. Voor de interpretatie van de tussenliggende waardes wordt conventioneel verwezen naar de beroemde vuistregels van statisticus Jacob Cohen, waarbij een *r* vanaf 0.10 een klein effect, een *r* vanaf 0.30 een middelgroot (*medium*), en vanaf 0.50 een groot effect representeert.[189] Hoewel deze vuistregels door een grote groep wetenschappers gedragen worden, is er – niet in de laatste plaats door Cohen zelf – ook kritiek op het feit dat ze maar al te vaak worden ingezet als vaststaande en contextafhankelijke empirische werkelijkheden, of zoals Joseph A. Durlak dat formuleert: 'as iron-clad criteria without reference to the measurements taken, the study design, or the practical [...] importance of the findings'.[190] Ook kleine effecten kunnen – afhankelijk van de context – bijzonder waardevol zijn. Durlak suggereert daarom altijd eerst te kijken naar eerder relevant onderzoek om een indicatie te krijgen van de omvang van de uitkomsten, en vooral Cohen's vuistregels te gebruiken wanneer dit onderzoek, deze context er niet is.[191]

Met de inzet van het classificatiemodel van Linders en Op de Beek is een dergelijke context er in deze scriptie er evident wel. Echter: waar Linders en Op de Beek laten zien hoe recensies significant verschillen tussen jaren en kranten, rapporteren ze geen effectgroottes. Om desalniettemin toch een indruk te krijgen van de reikwijdte van de verschillen die we kunnen verwachten, heb ik dezelfde vergelijkingen tussen jaren en kranten op mijn eigen dataset gedraaid, en daarvan de effectgroottes berekend. Meer bepaald heb ik eerst alleen de Nederlandse kranten in mijn dataset geselecteerd,[192] en vervolgens op paarsgewijze wijze *alle* kranten onderling en alle jaargangen op alle aspect- en eigenschapscodes vergeleken. Voor de 16 Nederlandse kranten in mijn dataset komt dit neer op 120 vergelijkingen en – met 27 labels in totaal – op 3.240 individuele statistische toetsen; voor de 73 jaren op 2.628 vergelijkingen en 70.956 statische toetsen. Om een verdere indruk te krijgen van de manieren waarop recensies van elkaar verschillen, heb ik eveneens de top 100 recensenten (met de meest recensies) onderling, de top 100 meest besproken schrijvers onderling, en de top 100 meest besproken boeken onderling met elkaar vergeleken. Dit komt voor elk van deze drie variabelen (recensent, schrijver, boek) neer op 4.950 vergelijkingen en 133.650 statistische toetsen.

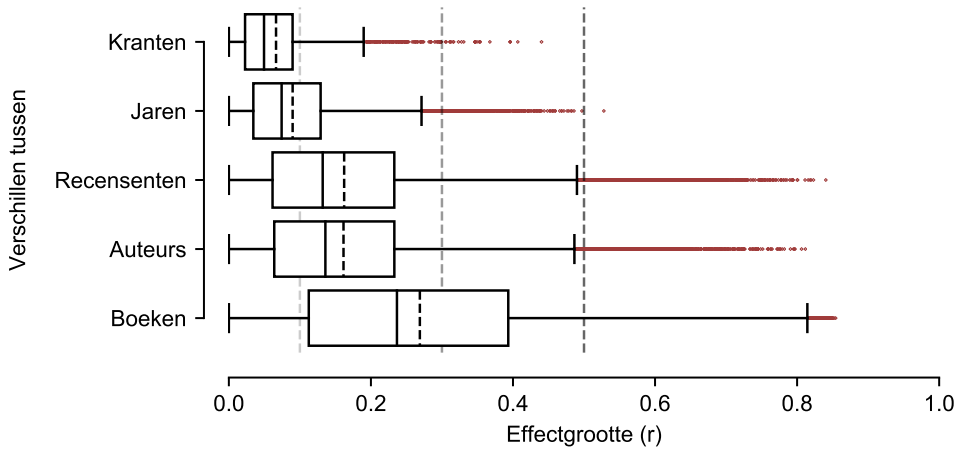
De boxplots in Figuur 6.1 geven een indruk van de distributie van effectgroottes over de vijf variabelen, waarin zowel significante als niet-significante verschillen zijn meegenomen. De figuur suggereert een relatie tussen de mate van specificiteit van de variabele en de effectgrootte. De effectgroottes van de vergelijkingen tussen kranten – misschien wel het meest algemene niveau – zijn relatief het kleinst: in termen van Cohen zouden alleen enkele uitschieters geclassificeerd kunnen als middelgroot. Dit geldt in zekere zin ook voor de verschillen tussen jaren. De reikwijdte van de effectgroottes van

189 J Cohen, *Statistical power analysis for the behavioral sciences. Abingdon*, 1988, 79-81.

190 Joseph A Durlak, "How to select, calculate, and interpret effect sizes", *Journal of pediatric psychology* 34, nummer 9 (2009): 922.

191 Ibidem, 922-923.

192 Dit voor de vergelijkbaarheid van de resultaten met de proefschriften van Linders en Op de Beek.



Figuur 6.1: Boxplots van de effectgroottes van alle onderlinge, paarsgewijze vergelijkingen op vijf variabelen. De gestreepte lijn in de doos geeft het gemiddelde aan; de gestreepte lijnen op de achtergrond markeren de grenswaardes die traditioneel worden aangehaald om kleine, middelgrote en grote effecten te duiden (0.1, 0.3 en 0.5).

de verschillen tussen recensenten en auteurs onderling is al een stuk groter, maar de grootste verschillen tussen recensies zijn simpelweg te herleiden naar verschillen tussen boeken. Met andere woorden: de aard van een recensie reflecteert tot op zekere hoogte de aard van de besproken roman. Dat is wellicht niet een bijzonder spannende of revolutionaire bevinding: als een roman een merkwaardig plot, een merkwaardige stijl of structuur heeft, als een roman uitgesproken (a)politiek of (a)religieus is, of vol zit met (al dan niet) geslaagde grappen, lijkt het mij – terechte nuances over de subjectiviteit van noties als 'merkwaardig' omwille van het argument daargelaten – ook zeer aannemelijk dat recensenten deze aspecten en eigenschappen zullen benoemen. Maar dit heeft dan wél als implicatie dat wanneer we groepen recensies op een minder specifiek niveau analyseren, we rekening moeten houden met het feit dat deze groepen onderling evenzeer – of zelfs: sterker – verschillen, en de effectgroottes dit ook in meer of minder mate zullen reflecteren.

Het lijkt mij op basis van de resultaten in Figuur 6.1 dan ook bijzonder onaannemelijk effectgroottes groter dan 0.5 in de vergelijking tussen Nederlandse en Vlaamse literatuurkritiek aan te zullen treffen. De herkomst van recensent representeert een niveau van abstractie vergelijkbaar met 'krant' of 'jaar van publicatie', en veel eerder verwacht ik dan ook effectgroottes binnen een dergelijke reikwijdte aan te zullen treffen, waarbij effectgroottes groter dan .20 al relatief zeldzaam zijn. En zonder elk minimaal verschil als betekenisvol te beschouwen, lijkt mij dat ook passen bij het vraagstuk van deze scriptie. Dat ik, zoals ik in de inleiding heb uiteengezet, verwacht dat Vlaams critici meer vanuit een politiek-poëticaal kader lezen dan Nederlandse critici, betekent niet dat ik verwacht dat dit verschil allesomvattend is – dat een Vlaams criticus die Herman Brusselmans in *De Morgen* dat onvermijdelijk politieker doet dan een Nederlands criticus Anja Meulenbelt in *De Waarheid*. Eerder verwacht ik, juist *ondanks* alle andere verschillen, onder

de streep toch een verschil tussen Nederlandse en Vlaamse literatuurkritiek aan te treffen.

Deel III

RESULTATEN

Hoofdstuk 7

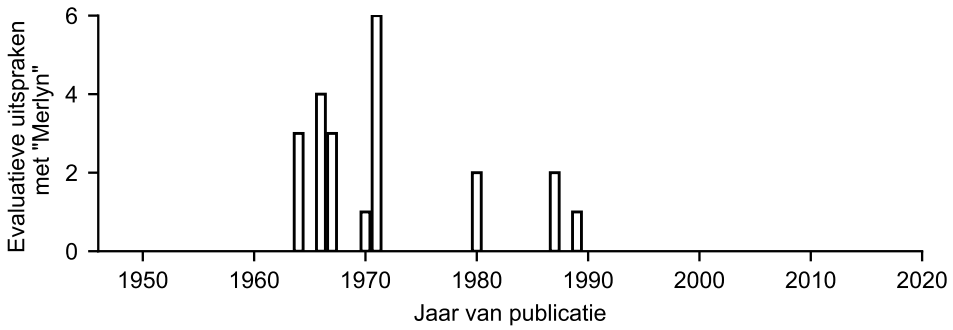
ASPECTEN IN NEDERLANDSE EN VLAAMSE RECENSIES

In 1989, ruim twintig jaar na de laatste uitgave van *Merlyn*, opent de Nederlandse criticus Willem Kuipers zijn bespreking van Herman Koch's *Red ons, Maria Montanelli* (1989) in de *Volkskrant* met een stilistische analyse van de eerste twee zinnen van de recent verschenen roman. Bijzonder noemenswaardig is die lezing niet – Kuipers meent dat de openingszin niet krachtig genoeg is, en stelt een alternatief voor. Maar dan vervolgt hij zelfbewust:

> Dit merlinistische geëmmer – *Merlyn* was een literair tijdschrift waarvan de medewerkers hun best deden verhalen en gedichten zeer nauwkeurig te lezen – moge voor lezers die het liefst in een zo hoog mogelijk tempo door de bladzijden razen, een hinderlijk gewicht aan de kritiek geven, het is onontkoombaar voor degenen die ook de verborgen kwaliteiten van een literair kunstwerk willen blootleggen. (24-02-1989, *Volkskrant*, W. Kuipers)

Hoewel het te betwijfelen is in hoeverre het aanwijzen van stilistische oneffenheden in de openingszin daadwerkelijk is te omschrijven als 'merlinistisch' (en zeker als 'zeer nauwkeurig'), is de tussenzin is het bovenstaand citaat veel belangwekkender. *Merlyn* is, zo lijkt het, in 1989 geen vanzelfsprekendheid meer, geen onderdeel van het gedeelde referentiekader tussen lezers en critici, maar een stuk geschiedenis – iets dat *was*, en een tussenzin met uitleg vereist. En hoewel Kuipers uiteindelijk wél pleit voor het nut van een dergelijke benadering, voor een bepaalde groep lezers althans, doet zijn typering van *Merlyn* ook enigszins banaal aan. De 'medewerkers' (niet de redacteuren, niet de critici) deden 'hun best' nauwkeurig te lezen – niet bepaald de toon die men zou verwachten bij een tijdschrift dat een grote stempel op de Nederlandse literatuurkritiek zou hebben gedrukt.

Deze verwijzing naar *Merlyn* is dan ook de laatste keer in het recensiecorpus dat het tijdschrift in een evaluatieve context wordt opgebracht (zie ook Figuur 7.1). Daarna wordt het tijdschrift nog vijf keer in een descriptieve context aangehaald, als onderdeel van een biografische beschrijving ('Jaap Oversteegen was redacteur van Merlyn'). Eveneens relevant in de context van deze scriptie is dat bijna alle verwijzingen naar *Merlyn* afkomstig zijn



Figuur 7.1: Het aantal evaluatieve uitspraken die het woord *Merlyn* (of varianten daarop) bevatten per jaar.

van Nederlandse critici: in Nederlandse recensies komt *Merlyn* (en varianten als merlinistisch en merlynse) 41 keer voor; in Vlaamse recensies slechts 1 keer. Dat verschil is (logischerwijs) significant ($\chi^2 = 8.53, p < 0.01$).

Dat *Merlyn* vanaf de jaren negentig niet meer in een evaluatieve context genoemd wordt, sluit tot op zekere hoogte aan bij de ideeën die onderzoekers als Kees van Rees, Jos Joosten, Nel van Dijk en Suzanne Janssen over de invloed van het tijdschrift hebben geformuleerd.[193] Deze zou zich aandienen vanaf de tweede helft van de jaren zestig, en in de loop van de jaren tachtig geleidelijk weer verdwijnen, waarna we vooral verscheidenheid in de literatuurkritiek zouden zien.[194] Maar een (enigszins zelfbevestigende) frequentieanalyse van het woord *Merlyn* volstaat natuurlijk allerminst om veranderende benaderingen in de Nederlandse literatuurkritiek te duiden – laat staan om daarmee het verschil tussen Nederlandse en Vlaamse recensies te karakteriseren. Dat een concrete bron langzaam uit het referentiekader van critici verdwijnt, of – in het geval van de Vlaamse recensies – volledig ontbreekt, betekent immers nog niet dat critici onverminderd hun (al dan niet onbewuste) zwaartepunten leggen op de werkimmanente aspecten van literaire werken.

Het analysemodel dat ik in het tweede deel van deze scriptie uitgebreid heb besproken, biedt wél de mogelijkheid om dergelijke subtiele accenten onder de loep te nemen. In dit hoofdstuk richt ik mij op het eerste deel van het analysemodel: de aspectcodes. Hoewel ik de resultaten voor elke aspectcode individueel rapporteer, ligt het zwaartepunt van mijn analyse op de verhouding en de ontwikkeling van werkimmanente aspecten tegenover meer buitenliteraire aspecten. Ik verwacht daarbij in Nederland, zoals ik in de inleiding ook deels uiteen heb gezet, (mede) onder invloed van *Merlyn* vanaf de jaren zestig een substantiële toename in werkimmanente

193 N van Dijk en S Janssen, "De reuzen voorbij. De metamorfose van de literaire kritiek in de pers sedert 1965", *J. Bardoel et al (red.), Journalistieke cultuur in Nederland, Amsterdam University Press, Amsterdam*, 2002, 209–235; CJ van Rees, "De brugfunctie van Merlyn in de Nederlandse literatuurbeschouwing", *K. Fens & H. Verdaasdonk (red.), Op eigen gronden. Opstellen aangeboden aan Prof. Dr. JJ Oversteegen (...), Utrecht*, 1989, 43–64; Jos Joosten, "Opkomst en ondergang van *Merlyn* in de Nederlandse literatuurkritiek", in *Staande receptie: over literatuur, kritiek en literatuurwetenschap* (Nijmegen: Vantilt, 2012), 59–68.

194 Joosten, "Opkomst en ondergang van *Merlyn* in de Nederlandse literatuurkritiek", 63; Dijk en Janssen, "De reuzen voorbij. De metamorfose van de literaire kritiek in de pers sedert 1965", 221.

Tabel 7.1: De resultaten van de aspectcodes over het gehele recensiecorpus. Gem. geeft de gemiddelde kans – uitgedrukt in percentages – voor elke aspectcode aan, SD de standaarddeviatie en de laatste drie kolommen zijn de statistieken van de Mann-Whitney toetsen met herkomst van de recensent als onafhankelijke variabele.

	Nederlands	Vlaams			
	Gem. (SD)	Gem. (SD)	U	p	r
Auteur/intentie	9.83 (7.19)	10.66 (7.09)	38676501.5	0.000	-0.06
Dialogen	1.16 (2.02)	1.08 (1.82)	44428897.0	0.000	0.04
Geheel/deel	23.39 (11.23)	24.61 (10.40)	39050989.0	0.000	-0.05
Overig	4.14 (3.95)	4.06 (3.62)	41787598.5	0.301	-0.01
Personages	8.81 (7.89)	8.22 (7.09)	43396297.0	0.002	0.02
Plot	8.81 (7.35)	7.74 (6.28)	45245568.0	0.000	0.05
Setting	4.91 (5.54)	4.83 (5.00)	40997006.5	0.002	-0.02
Stijl	17.49 (10.48)	16.86 (9.42)	43135719.5	0.018	0.02
Structuur	7.74 (5.99)	8.26 (5.83)	39330228.5	0.000	-0.05
Symboliek	1.02 (0.80)	0.99 (0.72)	42639103.0	0.264	0.01
Thematiek	10.36 (7.42)	10.40 (6.79)	41152512.0	0.008	-0.02
Uiterlijk en presentatie	2.34 (3.61)	2.28 (3.36)	40465768.0	0.000	-0.03

aspecten aan te treffen, en dan in het bijzonder van de structuuraspecten, terwijl buitenliteraire aspecten nog tot de jaren tachtig hun stempel op de Vlaamse recensies blijven drukken. Een laatste belangrijke kanttekening is daarbij op zijn plaats. Een frequentieanalyse van vermeldingen van *Merlyn* is wellicht oppervlakkig, maar een directe relatie tot het tijdschrift kunnen we er wel mee leggen. Dat is niet noodzakelijk het geval bij de aspectcodes: we kunnen nooit met zekerheid zeggen of een toename van structuurevaluaties toe te schrijven is aan de invloed van het tijdschrift, of bijvoorbeeld aan de ontzuiling. Een dergelijke invloed (sowieso een 'heikele categorie', in termen van Buelens[195]) acht ik in geen geval empirisch aantoonbaar. De geformuleerde verwachtingen moeten dan ook vooral gelezen worden als een onderbouwde manier om mogelijke verschillen te kunnen interpreteren en aannemelijk te maken.

7.1 SYNCHRONE ANALYSE

Tabel 7.1, waarin het gemiddelde aandeel van de aspectcodes in Nederlandse en Vlaamse recensies wordt gerapporteerd, toont dan ook direct een interessant contrast met de vergelijkende frequentieanalyse. Waar *Merlyn* significant vaker voorkomt in Nederlandse recensies, besteden Nederlandse recensenten – over het gehele corpus bezien – niet meer aandacht aan structuuraspecten dan hun Vlaamse collega's. Sterker nog: de kans dat een structuuraspect voorkomt in een recensie is juist hoger wanneer deze geschreven is door een Vlaming dan wanneer deze van Nederlandse hand is. Dit verschil is significant – al is de effectgrootte beperkt.

195 Buelens, *Van Ostaijen tot heden: zijn invloed op de Vlaamse poëzie*, 23.

Dat geldt overigens niet alleen voor de structuurevaluaties. Op de codes *overig* en *symboliek* na vinden we voor alle aspectcodes significante verschillen tussen Vlaamse en Nederlandse critici, hetgeen gezien de steekproefgrootte van deze studie ook enigszins in de lijn der verwachting lag (zie ook Hoofdstuk 6.3). Tegelijkertijd zijn de verschillen tussen de gemiddeldes klein. Het relatief grootste verschil treffen we aan bij het aspect *auteur/intentie*, dat in Vlaamse recensies vaker aan bod komt, met een effectgrootte (r) van -0.06. Dat is ongeveer even hoog als de *gemiddelde* effectgrootte die we aantreffen wanneer we krantentitels onderling vergelijken.[196] Anders geformuleerd: de verschillen tussen kranten (en daarmee ook: tussen jaren, recensenten, boeken en auteurs) zijn, over het hele corpus genomen, veel groter dan die tussen Vlaamse en Nederlandse critici.

De (beperkte) verschillen die we aantreffen, laten daarbij ook niet direct een eenduidig beeld zien. In Vlaamse recensies zijn er weliswaar meer structuurevaluaties te vinden – bij uitstek dé categorie die het 'meest verbonden is aan de poëticale opvattingen *Merlyn*', aldus Op de Beek[197] – maar ook, zoals gezegd, meer auteursevaluaties, waarin de kritiek nadrukkelijk niet werkgericht is. Bovendien komen de andere meer werkimmanente aspecten – *stijl*, *plot*, *dialogen* en *personages* – juist vaker voor in Nederlandse recensies, en komen er in Vlaamse recensies – al is het verschil buitengewoon klein – significant meer thematiekevaluaties voor, waarin de relatie met de wereld buiten de roman al veel sneller kan worden gelegd.[198] Een omgekeerde hypothese, dat Vlaamse critici juist sterker autonoom-esthetisch zouden beoordelen, kan dus eveneens niet zozeer ondersteund worden. Eerder zouden we, wanneer we de verschillen bij elkaar optellen, voorzichtig uit Tabel 7.1 kunnen concluderen dat Nederlandse critici, ondanks dat structuurevaluaties vaker voorkomen in Vlaamse recensies, inderdaad vanuit een meer werkgericht kader beoordelen – zij het in beperkte mate.

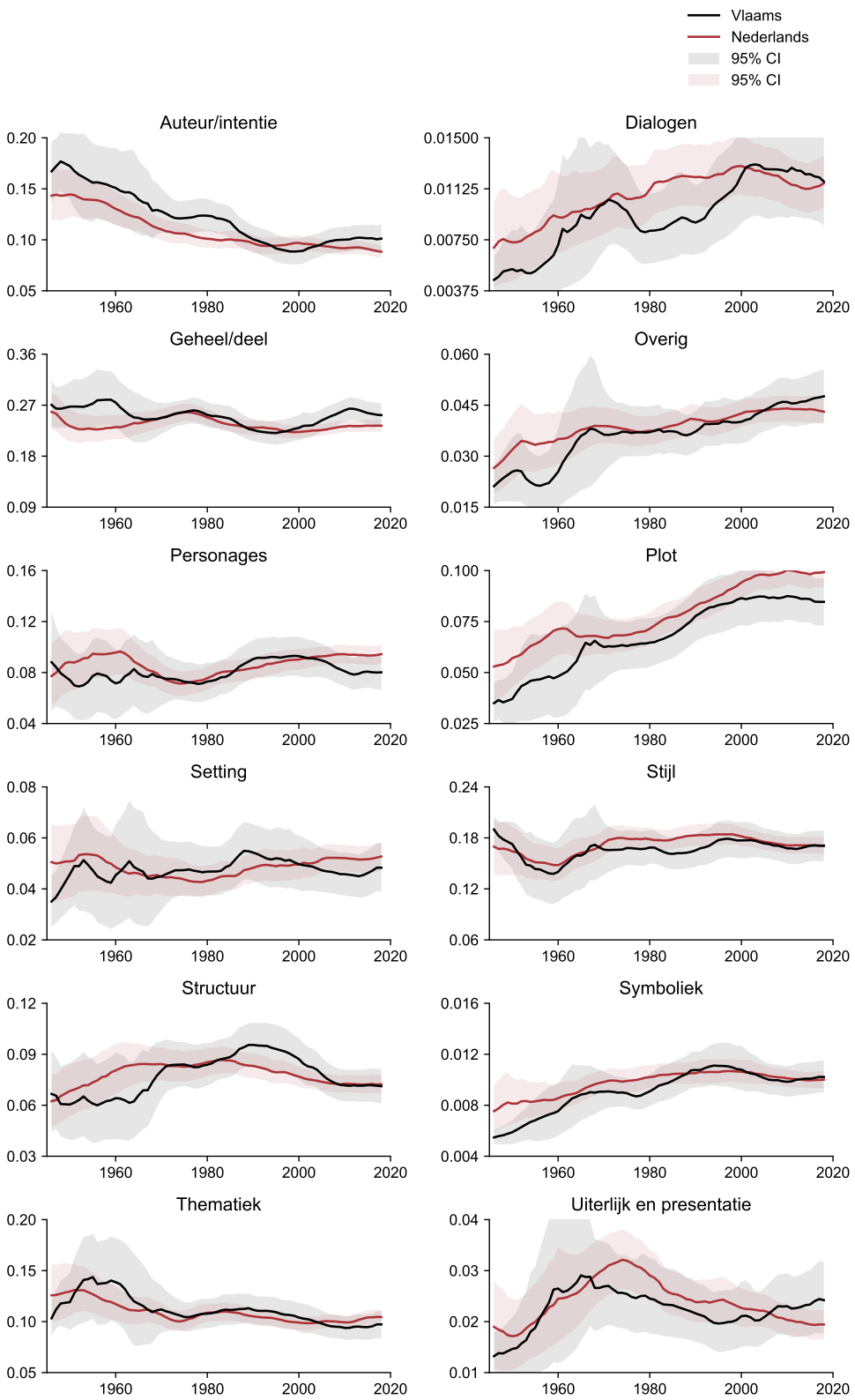
7.2 DIACHRONE ANALYSE

Wanneer we kijken naar de manier waarop de voorkeur voor bepaalde aspecten in de Nederlandse en Vlaamse kritiek door de tijd heen is veranderd, kunnen we al grotere verschillen – zowel tussen jaren als tussen Nederlandse en Vlaamse critici – waarnemen (zie Figuur 7.2). Zo is het aandeel van het aspect *auteur/intentie* in Vlaamse recensies tussen 1946 en 2018 bijna gehalveerd. In Nederland zien we een gelijkaardige gestaag dalende tendens, al was het aandeel van auteursevaluaties daar aanvankelijk al kleiner. Tegenover deze halvering staat de verdubbeling van het aspect *plot*, die we zowel in de Nederlandse als in de Vlaamse literatuurkritiek aantreffen, maar waarbij plotevaluaties in Nederlandse recensies wel structureel vaker voorkomen dan in Vlaamse recensies.

196 Die is namelijk 0.066. Zie ook Figuur 6.1.

197 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 240.

198 Zie Linders, "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005", 206-207.



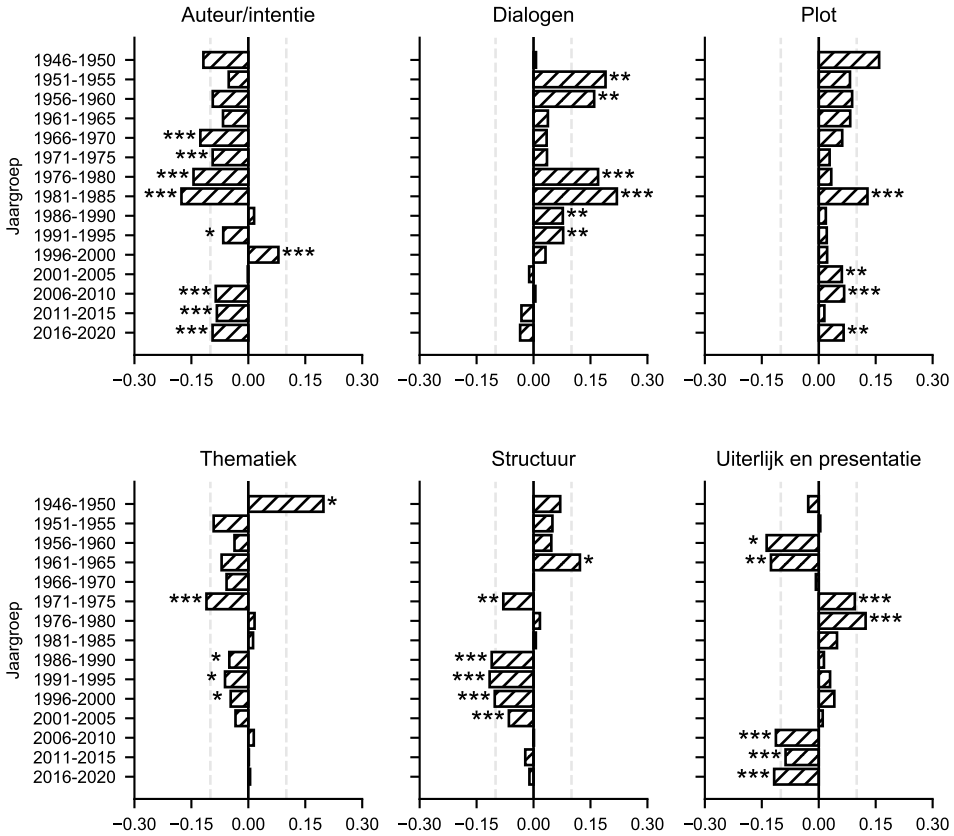
Figuur 7.2: De proportionele aandacht voor de aspectcodes door de tijd heen. Omdat, wanneer ik de Y-assen gelijk zou houden, het gevaar bestaat dat relatief betekenisvolle ontwikkelingen van laagfrequente labels niet zichtbaar zijn (en andersom), heb ik de assen relatief geschaald, zodat elk ontwikkeling van de onderkant van de Y-as naar de bovenkant voor elk label een verviervoudig markeert (waarbij ik steeds de kleinste verviervoudiging heb geselecteerd). De lijndiagrammen moeten daarom ook niet zozeer in vergelijking met elkaar gelezen worden. Zie Tabel 7.1 voor de onderlinge verhoudingen.

Ook het aandeel van het aspect *dialogen* is zowel in Nederlandse als in Vlaamse recensies meer dan verdubbeld. In recensies van Nederlandse critici verloopt deze ontwikkeling redelijk rechtlijnig; in Vlaamse recensies verloopt ze iets grilliger. Ondanks dat de aandacht voor *dialogen* is toegenomen, blijft het aandeel ervan beperkt. In recensies uit 2018 is *dialogen* nog steeds een van de minst besproken aspecten – alleen het aandeel van *symboliek* is beperkter. Het concrete effect van de verdubbeling van *dialogen* op de aard van de recensies is daarom enigszins beperkt te noemen. Dat geldt in zekere zin ook voor het aandeel van het aspect *uiterlijke en presentatie*, dat in Nederlandse recensies tussen 1946 en 1980, en in Vlaamse recensies zelfs tussen 1946 en 1970, is verdubbeld, en daarna weer vrijwel even snel is afgenomen.

Het aandeel van het aspect *thematiek* is sinds 1946 zowel in Vlaamse als Nederlandse recensies lichtelijke afgenomen. Maar anders dan bijvoorbeeld bij *auteur/intentie* voltrekt deze ontwikkeling zich in Nederland en Vlaanderen niet op een gelijkaardige manier. Waar Nederlandse critici vanaf 1946 consistent minder aandacht hebben voor *thematiek* (een beperkte toename rond 1975 uitgezonderd), neemt het aandeel van thematiekevaluaties in prozarecensies van Vlaamse critici pas midden jaren zestig af, en lijkt er eerst nog een stijging waarneembaar.

In de ontwikkeling van het aspect *structuur* zijn eveneens verschillen tussen Vlaamse en Nederlandse recensies te zien. Waar uit de vorige paragraaf bleek dat Vlaamse recensies over het hele corpus bekeken juist meer structuurevaluaties bevatten, laat een longitudinale analyse zien dat voor ongeveer 1975 de verhoudingen juist andersom liggen. In Nederland nemen structuurevaluaties sinds 1946 vrij consequent toe. Deze toename voltrekt zich in de eerste twee decennia relatief snel. Daarna neemt de aandacht voor structuur in Nederlandse recensies nog in beperkte mate toe, tot 1983 – het jaar waarin het relatieve aandeel van het aspect *structuur* het hoogst is. In Vlaamse recensies, daarentegen, blijft het aandeel van structuurevaluaties in de eerste twee decennia na 1946 opmerkelijk gelijk. Daar nemen de structuurevaluaties pas vanaf midden jaren zestig toe, waarbij we ruwweg twee fases van substantiële stijging kunnen onderscheiden. De eerste voltrekt zich vanaf midden jaren zestig tot begin jaren tachtig, waarna het aandeel van structuurevaluaties ongeveer in Vlaamse recensies ongeveer gelijk is aan het aandeel in Nederlandse recensies. Maar waar er in Nederland na 1983 juist weer een lichte kentering waar te nemen is, neemt het aandeel van structuurevaluaties in Vlaanderen juist voor een tweede keer toe, tot midden jaren 1990.

De aandacht voor het aspect *personages* neemt, in tegenstelling tot *structuur*, in Nederlandse recensies rond de jaren zestig juist sterk af, om vanaf de tweede helft van de jaren zeventig weer lichtelijke toe te nemen. In Vlaamse recensies daarentegen blijft het aandeel van *personages* tot de jaren tachtig vrij constant. Tussen de jaren tachtig en negentig neemt het aandeel van *personages* relatief snel toe, tot het begin van 2000, waarna het weer gestaag afneemt. De overige aspecten – *setting*, *stijl*, *geheel/deel* en *symboliek* – verlopen, zoals Linders eveneens al voor Nederlandse recensies heeft gerapporteerd, rela-



Figuur 7.3: De vijfjaarlijkse effectgroottes (r) van *auteur/intentie*, *dialogen*, *plot*, *thematiek*, *structuur* en *uiterlijk en presentatie*. Een positieve waarde geeft aan dat een aspect in Nederlandse recensies een hoger aandeel heeft; een negatieve waarde geeft aan dat een aspect in Vlaamse recensies een hoger aandeel heeft. *** $p < 0.001$; ** $p < .01$; * $p < .05$

tief constant.[199] Daarbij laten deze aspecten, op enkele uitschieters na, geen duidelijke verschillen tussen Nederlandse en Vlaamse literatuurkritiek zien.

Het zijn dus vooral de aspecten *auteur/intentie*, *dialogen*, *plot*, *thematiek*, *structuur*, en *uiterlijk en presentatie* waarbij we – op specifieke momenten in de tijd althans – noemenswaardige verschillen tussen Vlaamse en Nederlandse recensies lijken aan te kunnen treffen. Ik schrijf hier 'lijken', omdat relatief grote verschillen tussen gemiddeldes nog niet noodzakelijk noemenswaardig zijn wanneer we niet ook de variantie van de data in onze analyse verdisconteren – zeker wanneer het aantal recensies beperkt is, zoals voor 1960 het geval is. Daarom voer ik op de zes bovengenoemde aspecten aanvullende toetsen uit. Hierbij allereerst heb ik het recensiecorpus onderverdeeld in periodes van vijf jaar, om vervolgens voor elke periode de verschillen tussen Nederlandse en Vlaamse recensies apart te toetsen. Figuur 7.3 toont voor de zes aspecten de effectgroottes en de bijbehorende p-waardes van de Mann Whitney-toetsen op de vijftien verschillende periodes.

Hoewel we ook hier geen effectgroottes groter dan .30 aantreffen, zijn de effectgroottes al aanmerkelijk groter dan de effectgroottes die we von-

199 Linders, "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005", 154.

den wanneer we Vlaamse en Nederlandse literatuurcritici over de gehele periode met elkaar vergeleken. In tegenstelling tot in Tabel 7.1 zijn er in dit figuur namelijk wel (significante) effectgroottes groter dan .10 te vinden. Dergelijke effectgroottes worden traditioneel als klein geduid worden, maar kunnen in vergelijking met de verschillen die we tussen kranten en jaren aantreffen al als relatief uitzonderlijk worden beschouwd. Daarbij zijn zowel voor *auteur/intentie*, *dialogen* en *plot* de grootste significante verschillen tussen Nederlandse en Vlaamse critici in 1981-1985 aan te treffen. Verder is het grootste verschil voor het aspect *thematiek* te vinden in 1946-1950, voor *structuur* in 1961-1965 en voor *uiterlijk en presentatie* in 1956-1960.

In de context van de al dan niet veranderende aandacht voor werkimmanente aspecten is vooral ook de (gedeeltelijke) symbiose van de effectgroottes van *auteur/intentie* en *structuur* betekenisvol. Auteursevaluaties komen tot 1985 structureel vaker voor in Vlaamse recensies dan in Nederlandse recensies. Na 1985 veranderen deze verhoudingen abrupt. In 1986-1990 is er geen significant verschil meer tussen Nederlandse en Vlaamse critici, in 1991-1995 zijn auteursevaluaties nog in slechts beperkte mate vaker in Vlaamse recensies te vinden, en in 1996-2000 zijn de verhoudingen zelfs omgekeerd, en kunnen we in Nederlandse recensies juist significant meer auteursevaluaties vinden. En waar de verschillen tussen Vlaamse en Nederlandse critici op het vlak van auteursevaluaties na 1985 juist diffuser worden, zien we in precies dezelfde periode de verschillen in het gebruik van het aspect *structuur* juist meer geprononceerd worden, en vinden we in Vlaamse recensies in 1986-1990, 1990-1995, 1996-2000 én 2001-2005 significant meer structuurevaluaties dan in Nederlandse recensies. Kijkend naar de auteurs- en structuurevaluaties kunnen we dus inderdaad bevestigen dat de Vlaamse literaire praktijk in 1985, zoals Hugo Brems heeft betoogd, op een belangrijk kruispunt stond.[200]

Hoewel het aandeel van structuurevaluaties in Nederlandse recensies in 1961-1965 significant hoger is dan in Vlaamse recensies, blijft het verwachte *Merlyn*-effect in Nederlandse recensies tot op zekere hoogte uit. Het aandeel van structuurevaluaties in Nederlandse recensies is tussen 1946 en 1983 weliswaar met 44% procent toegenomen, maar het grootste deel van de toename vindt plaats vóór de eerste publicatie van *Merlyn*.[201] De resultaten suggereren dan ook eerder een algemeen (vroeg) ontzuilingseffect dan een specifieker *Merlyn*-effect, waarmee *Merlyn* ook meer, zoals Jos Joosten heeft beargumenteerd, gelezen moet worden als 'een symptoom van de ontzuiling'.[202]

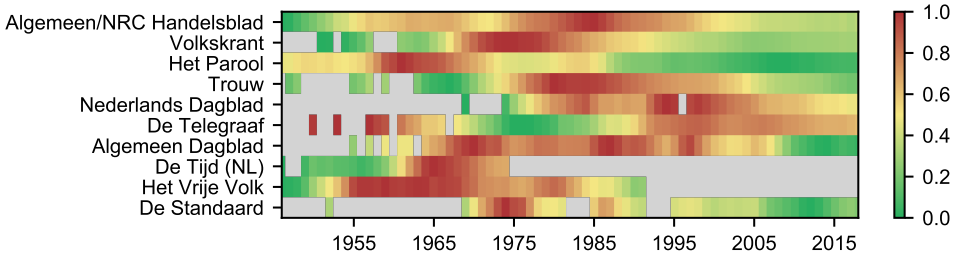
Het ontbreken van een substantieel *Merlyn*-effect contrasteert met de bevindingen van Linders en Op de Beek. Ze rapporteren weliswaar dat de aandacht voor structuurevaluaties het snelst stijgt tussen 1955 en 1965,[203] maar uit hun analyse komt eveneens naar voren dat structuurevaluaties tus-

200 Brems, *Altijd weer vogels die nesten beginnen: geschiedenis van de Nederlandse literatuur, 1945-2005*, 486.

201 Tussen 1946 en 1961 is het aandeel van structuurevaluaties al met 35% toegenomen. Tussen 1961 en 1983 neemt het aandeel van structuurevaluaties nog met 6,4% toe.

202 Joosten, "Opkomst en ondergang van *Merlyn* in de Nederlandse literatuurkritiek", 62.

203 Dit resultaat wordt overigens nauwelijks geduid, omdat Linders en Op de Beek verschillen tussen jaren als zijnde nominale variabelen paarsgewijs toetsen en 1965 zodoende niet als een significant jaar naar voren komt.



Figuur 7.4: De relatieve aandacht voor structuurevaluaties door de jaren heen per krant. De kranten zijn geordend op basis van het totaal aantal recensies in het recensiecorpus. Hier wordt alleen de top tien kranten weergeven.

sen 1965 en 1985 met bijna 51% zouden zijn toegenomen – een resultaat dat al veel meer toe te schrijven is aan de erfenis van *Merlyn*.

Hoe is dit verschil te verklaren? Allereerst kunnen we natuurlijk niet geheel uitsluiten dat een deel van het verschil met de resultaten van Linders en Op de Beek het toevallige gevolg is van classificatiefouten van het automatische meetinstrument. Tegelijkertijd telt het corpus van deze scriptie meer dan 2000 Nederlandse recensies tussen 1975 en 1985 (de periode waarin we voornamelijk verschillen met Linders en Op de Beek aantreffen), waardoor de invloed van individuele classificatiefouten in principe beperkt zou moeten zijn. De meest aannemelijke verklaring lijkt mij dan ook eerder het verschil in corpusselectie. Waar Linders en Op de Beek zich hebben beperkt tot vijf dagbladen, heb ik elk nationaal dagblad waarvan er in *LiteRom* recensies beschikbaar zijn in mijn corpus opgenomen.

Om deze veronderstelling verder te exploreren heb ik de aandacht voor structuurevaluaties door de jaren heen uitgesplitst per krant (zie Figuur 7.4). De figuur toont, per krant, de *relatieve* aandacht voor structuurevaluaties van 1946 tot 2019. Relatief betekent hier dat de kleuren alleen markeren hoe groot het aandeel van structuurevaluaties is vergeleken met andere jaren van dezelfde krant. Zodoende kunnen we zien dat de proportionele aandacht voor structuurevaluaties in de kranten die opgenomen zijn in het corpus van Linders en Op de Beek – *Algemeen Dagblad, Algemeen/NRC Handelsblad, Het Parool, Trouw* en *de Volkskrant* – over het algemeen piekt tussen 1970 en 1985. Alleen *Het Parool*, waar al een decennium eerder de meeste aandacht naar het aspect *structuur* uitgaat, vormt hierop een uitzondering. Van de Nederlandse kranten die *niet* zijn opgenomen in de selectie van Linders en Op de Beek, daarentegen, piekt alleen *Nederlands Dagblad* in ongeveer dezelfde periode. Bij *Het Vrije Volk* en *De Tijd* – waar, zoals gezegd, Kees Fens recensies schreef voor hij naar de *Volkskrant* vertrok[204] – zien we de proportionele aandacht voor structuurevaluaties al rond de jaren zestig pieken, hetgeen inderdaad een afvlakkend effect heeft op de toename tussen 1955 en 1985. De aandacht voor structuurevaluaties verloopt in *De Telegraaf* bovendien vrijwel tegenovergesteld aan de algemene trend. Daar neemt het aandeel van structuurevaluaties tussen 1960 en 1985 juist af, om weer toe te nemen op het moment dat het aspect *structuur* in andere kranten juist weer minder wordt besproken. Ook dit zou verklaren waarom de toename van structuur-

204 Zie Hoofdstuk 5.1.1.

evaluaties tussen 1975 en 1985 in deze scriptie veel minder geprononceerd is.

7.3 VERSCHILLEN TUSSEN VLAAMSE EN NEDERLANDS SCHRIJVERS

Waar ik mij toe nu toe heb geconcentreerd op verschillen tussen Nederlandse en Vlaamse recensies als geheel, kunnen verschillen tussen Noord en Zuid zich ook *binnen* deze groepen manifesteren, namelijk op het niveau van de besproken schrijver. Zo hebben onder andere Hugo Brems en Kevin Absillis betoogd dat Vlaamse auteurs in de Nederlandse literatuurkritiek veelal expliciet als 'Vlaams', als 'anders' worden gelezen.[205] Vanuit dit perspectief is niet alleen relevant om te onderzoeken in hoeverre Vlaamse en Nederlandse schrijvers anders beoordeeld worden door Vlaamse en Nederlandse critici, in hoeverre er dus bij Vlaams en Nederlandse schrijvers andere aspecten aan bod komen, maar ook: in hoeverre de verschillen die we in de vorige twee paragrafen hebben geïdentificeerd ook gelden wanneer we alleen de recensies van Vlaamse schrijvers – of alleen Nederlandse schrijvers – analyseren.

Om dit te achterhalen heb ik het corpus opgedeeld in vier groepen: Nederlandse critici die romans van Nederlandse auteurs bespreken, Nederlandse critici die Vlaamse auteurs bespreken, Vlaamse critici die Nederlandse auteurs bespreken en Vlaamse critici die Vlaamse auteurs bespreken.[206] Daarbij heb ik allereerst voor elk aspect aan de hand van een Kruskal-Wallistoets bepaald of er een hoofdeffect voor deze groepen is. Dit bleek het geval voor alle aspecten behalve *overig* en *symboliek*.[207] Vervolgens heb ik, voor alle aspecten waarvan er sprake is van een hoofdeffect, de vier groepen paarsgewijs vergeleken aan de hand van Mann-Whitneytoetsen. Tabel 7.2 toont het resultaat van deze paarsgewijze vergelijkingen.[208]

---

205 Kevin Absillis, "'From now on we speak civilized Dutch': the authors of Flanders, the language of the Netherlands, and the readers of A. Manteau", *Language and Literature* 18, nummer 3 (2009): 265–280; Hugo Brems, "Het kritische discours over Vlaamse literatuur in Nederlandse kranten en weekbladen (1980-1994)", *Nederlandse Letterkunde* 4 (1999): 98–115.

206 Hierbij heb ik, mede om de omvang van deze scriptie enigszins beperkt te houden, niet meegenomen of deze auteurs in Nederland of in Vlaanderen of in Nederland publiceerden. Dat zou echter ook betekenisvol kunnen zijn. Zo spreekt Ton Anbeek over drie verschillende Nederlandstalige literaturen: 'de Nederlandse, de Vlaamse, en de Vlaamse voorzover erkend door de Nederlandse uitgevers', en corrigeert Geert Buelens hem terecht door te stellen dat er in dat geval vier literaturen zouden moeten zijn: 'de Nederlandse, de Vlaamse, de Vlaamse voor zover bekend in Nederland en de Nederlandse voor zover die doordringt tot Vlaanderen'. Een interessant vervolgonderzoek zou, alleen al op basis van de metadata, kunnen analyseren in hoeverre de vier groepen die ik in deze paragraaf onderscheid overstemmen met deze vier literaturen. Zie Anbeek, *Geschiedenis van de Nederlandse literatuur tussen 1885 en 1985*, 17; Buelens, "Gotspe of godsgeschenk? Enkele notities over zin en onzin van een Belgische literatuurgeschiedenis", 71.

207 *Auteur/intentie*: H(4)=110.94, p=<.001; *Dialogen*: H(4)=42.56, p=<.001; *Geheel/deel*: H(4)=75.87, p=<.001; *Overig*: H(4)=7.56, p=0.056; *Personages*: H(4)=35.95, p=<.001; *Plot*: H(4)=74.02, p=<.001; *Setting*: H(4)=11.02, p=<.05; *Stijl*: H(4)=10.37, p=<.05; *Structuur*: H(4)=54.54, p=<.001; *Symboliek*: H(4)=2.73, p=0.435; *Thematiek*: H(4)=8.98, p=<.05; *Uiterlijk en presentatie*: H(4)=42.93, p=<.001

208 Bij de Mann-Whitneytoetsen is gecorrigeerd voor herhaalde toetsen aan de hand van Bonferroni-correctie, waarbij een oorspronkelijk significatieniveau van 0.05 is aangehouden.

Tabel 7.2: Effectgroottes (r) van de paarsgewijze vergelijkingen waarbij groepen niet zijn vergeleken aan de hand van de herkomst van de recensent, maar ook de herkomst van de auteur.

Label	Groep	Gem.	SD	1	2	3	4
Auteur/intentie	1. Ned. criticus, Ned. auteur	9.81	7.19	-			
	2. Ned. criticus, Vl. auteur	10.01	7.18	ns	-		
	3. Vl. criticus, Ned. auteur	10.11	6.84	0.02	ns	-	
	4. Vl. criticus, Vl. auteur	11.21	7.28	0.08	0.09	0.08	-
Dialogen	1. Ned. criticus, Ned. auteur	1.17	2.03	-			
	2. Ned. criticus, Vl. auteur.	1.10	1.91	ns	-		
	3. Vl. criticus, Ned. auteur	1.07	1.84	-0.02	ns	-	
	4. Vl. criticus, Vl. auteur	1.09	1.80	-0.04	ns	ns	-
Geheel/deel	1. Ned. criticus, Ned. auteur	23.37	11.23	-			
	2. Ned. criticus, Vl. auteur	23.55	11.25	ns	-		
	3. Vl. criticus, Ned. auteur	24.09	10.42	0.03	ns	-	
	4. Vl. criticus, Vl. auteur	25.12	10.36	0.06	0.08	0.05	-
Overig	1. Ned. criticus, Ned. auteur	4.12	3.96	-			
	2. Ned. criticus, Vl. auteur	4.31	3.92	ns	-		
	3. Vl. criticus, Ned. auteur	4.13	3.54	ns	ns	-	
	4. Vl. criticus, Vl. auteur	3.99	3.70	ns	ns	ns	-
Personages	1. Ned. criticus, Ned. auteur	8.86	7.92	-			
	2. Ned. criticus, Vl. auteur	8.48	7.64	ns	-		
	3. Vl. criticus, Ned. auteur	8.76	7.36	ns	ns	-	
	4. Vl. criticus, Vl. auteur	7.70	6.77	-0.04	ns	-0.07	-
Plot	1. Ned. criticus, Ned. auteur	8.87	7.43	-			
	2. Ned. criticus, Vl. auteur	8.27	6.65	ns	-		
	3. Vl. criticus, Ned. auteur	8.06	6.56	-0.03	ns	-	
	4. Vl. criticus, Vl. auteur	7.43	5.97	-0.06	-0.06	-0.04	-
Setting	1. Ned. criticus, Ned. auteur	4.93	5.58	-			
	2. Ned. criticus, Vl. auteur	4.76	5.23	ns	-		
	3. Vl. criticus, Ned. auteur	4.90	5.37	ns	ns	-	
	4. Vl. criticus, Vl. auteur	4.77	4.59	0.02	0.04	ns	-
Stijl	1. Ned. criticus, Ned. auteur	17.45	10.47	-			
	2. Ned. criticus, Vl. auteur	17.77	10.61	ns	-		
	3. Vl. criticus, Ned. auteur	17.15	9.79	ns	ns	-	
	4. Vl. criticus, Vl. auteur	16.57	9.04	-0.02	-0.05	ns	-
Structuur	1. Ned. criticus, Ned. auteur	7.73	6.00	-			
	2. Ned. criticus, Vl. auteur	7.80	5.95	ns	-		
	3. Vl. criticus, Ned. auteur	8.25	5.96	0.03	0.04	-	
	4. Vl. criticus, Vl. auteur	8.28	5.70	0.05	0.06	ns	-
Symboliek	1. Ned. criticus, Ned. auteur	1.01	0.80	-			
	2. Ned. criticus, Vl. auteur	1.03	0.76	ns	-		
	3. Vl. criticus, Ned. auteur	0.99	0.68	ns	ns	-	
	4. Vl. criticus, Vl. auteur	1.00	0.75	ns	ns	ns	-
Thematiek	1. Ned. criticus, Ned. auteur	10.34	7.41	-			
	2. Ned. criticus, Vl. auteur	10.56	7.55	ns	-		
	3. Vl. criticus, Ned. auteur	10.31	6.69	ns	ns	-	
	4. Vl. criticus, Vl. auteur	10.48	6.89	0.02	ns	ns	-
Uiterlijk en presentatie	1. Ned. criticus, Ned. auteur	2.34	3.59	-			
	2. Ned. criticus, Vl. auteur	2.36	3.82	ns	-		
	3. Vl. criticus, Ned. auteur	2.19	3.39	ns	ns	-	
	4. Vl. criticus, Vl. auteur	2.36	3.32	0.05	0.07	0.07	-

ns = niet significant, Ned. = Nederlands, Vl. = Vlaams

Verrassend genoeg zijn er in Nederlandse recensies geen significante verschillen tussen de manier waarop Vlaamse auteurs en Nederlandse auteurs besproken worden. Ook *stijl* – het aspect waarvan Vlaamse literatuur wellicht nog het meest evident af kan wijken van de Nederlandse literatuur – wordt niet significant vaker besproken in recensies van romans van Vlaamse schrijvers. In Vlaamse recensies zien we daarentegen wél verschillen tussen besprekingen van Vlaamse schrijvers en Nederlandse schrijvers. Vlaamse critici hebben bij Nederlandse auteurs significant meer aandacht voor de aspecten *personages* en *plot*, terwijl ze bij Vlaamse auteurs juist meer aandacht hebben voor *auteur/intentie*, *geheel/deel*, en *uiterlijk en presentatie*. Vooral *auteur/intentie* is daarbij belangwekkend. Niet alleen vinden we daar, over alle aspecten bekeken, de grootste verschillen tussen de vier groepen, ook blijkt de herkomst van de auteur – in Vlaamse recensies althans – net zo belangrijk als de herkomst van de recensent voor de mate waarin auteursevaluaties voorkomen. Zo is het verschil in het aandeel van *Auteur/intentie* tussen Vlaamse recensies ongeveer even groot als het verschil tussen Nederlandse en Vlaamse recensies wanneer deze Vlaamse schrijvers bespreken. Wanneer we bovendien alleen kijken naar de recensies van het werk van Nederlandse schrijvers, zijn de verschillen tussen Vlaamse en Nederlandse critici aanmerkelijk kleiner.

Hoewel we in Nederlandse recensies geen verschillen tussen Vlaamse en Nederlandse romans aantreffen, vinden we wel meer verschillen *tussen* Nederlandse en Vlaamse recensies wanneer deze Vlaamse schrijvers bespreken dan wanneer ze Nederlandse schrijvers bespreken. In recensies van Vlaamse romans verschillen Nederlandse en Vlaamse critici op zeven van de twaalf aspecten (*auteur/intentie, geheel/deel, plot, setting, stijl, structuur, uiterlijk en presentatie*); in recensies van Nederlandse romans op vijf van de twaalf (*auteur/intentie, geheel/deel, dialogen, plot, structuur*). Daarbij zijn de verschillen tussen Nederlandse en Vlaamse critici structureel kleiner in recensies van Nederlandse romans dan in recensies van Vlaamse romans. Desalniettemin kunnen we wel concluderen dat de belangrijkste verschillen die we in de vorige twee paragrafen hebben geïdentificeerd – vooral op de aspecten *auteur/intentie*, *plot*, *structuur* – ook gelden wanneer alleen besprekingen van Nederlandse auteurs, of alleen besprekingen van Vlaamse auteurs, geanalyseerd worden.

Hoofdstuk 8

Eigenschappen in Nederlandse en Vlaamse recensies

In het vorige hoofdstuk vergeleken we de aspecten van literaire werken die Vlaamse en Nederlandse critici bespreken in dagbladrecensies. Zo blijkt er onder meer in Vlaamse recensies structureel meer aandacht uit te gaan naar de auteur van de roman, terwijl het plot voor Nederlandse recensenten een belangrijker aspect vormt. Ook zagen we hoe structuurevaluaties in de jaren zestig in Nederlandse recensies vaker voorkomen, terwijl er in de jaren tachtig en negentig juist in Vlaamse recensies meer aandacht wordt besteed aan de structuur.

Maar de aspecten alleen vertellen ons niet het volledige verhaal. Een toename in structuurevaluaties suggereert weliswaar dat literaire aspecten steeds meer de boventoon gaan voeren, het betekent niet noodzakelijk dat ook alleen nog de zuiver esthetische waarde van deze literaire aspecten aan bod komt. Ook werkimmanente aspecten kunnen politiek geëvalueerd worden. Neem *Max Havelaar*, waarover de meeste twintigste-eeuwse critici – zelfs Sötemann, wiens *De structuur van Max Havelaar* het schoolvoorbeeld van formalistische analyse vormt – het eens zijn dat de gelaagde vertelstructuur vooral in dienst staat van de politieke en morele aanklacht.[209] Of neem de narratologische benadering van onder andere Mieke Bal, waarin juist formele aspecten van literaire werken ingezet worden om uitspraken te doen over achterliggende ideologieën.[210]

Ook in het recensiecorpus zijn tal van evaluatieve uitspraken te vinden waarin een formele literaire eigenschap wordt verbonden met de maatschappelijke dimensie van een roman. Zo probeert Jos Borré in een recensie van *Het verdriet van België* de structuur van de dan (1983) recent verschenen roman te ontleden. Het is 'een kompleks boek', schrijft hij, '[m]en kan er geen

209 Saskia Pieterse, "De buik van de lezer. Over spreken en schrijven in Multatuli's Ideeën", 2008, 266.

210 Roel Smeets en andere, "Character Centrality in Present-Day Dutch Literary Fiction", *DH Benelux journal*, 2019, 73; Mieke Bal, "Narratology: Introduction to the Theory of Narrative", 2009,

onderdelen uit afzonderen zonder het geheel tekort te doen, terwijl het geheel op zich onoverzichtelijk veelzijdig is'. Borré vervolgt:

> Het evolueert traag over een zeer breed spoor, alsof Claus het verloop van dat stuk geschiedenis stil wilde leggen om nauwgezet de mozaïek samen te stellen van het samengaan van een bepaalde Vlaamse mentaliteit en bepaalde politiek-ideologische omstandigheden, een samengaan dat een verpletterende, verstikkende invloed moest hebben op de generatie van jongeren die erdoor werden meegesleept en nadien weer uitgespuwd. (12-03-1983, De Morgen, Jos Borré)

De structuur van het boek is complex, onoverzichtelijk, traag – niet bepaald termen die traditioneel positief geëvalueerd worden – maar die complexiteit, die onoverzichtelijkheid, die traagheid heeft voor Borré functie in de context van de maatschappelijke-politieke boodschap van de roman, en daarmee krijgen de formele aspecten – de structuur, het tempo – dus ook politieke eigenschappen toegewezen. Een analyse van alleen de aspecten zou hier dus een vertekend beeld opleveren: de aspecten zijn weliswaar ergocentrisch, de eigenschappen die aan deze aspecten worden toegekend zijn dat allerminst.

In dit hoofdstuk richt ik mij dan ook op het tweede deel van het analysemodel van Linders en Op de Beek: de eigenschapscodes. Net als in het voorgaande hoofdstuk rapporteer ik de resultaten voor elke eigenschapcode individueel, maar ligt het zwaartepunt van mijn analyse op de verhouding en de ontwikkeling van de meer werkimmanente eigenschappen – *continuïteit*, *helderheid*, *samenhang* – tegenover de 'ideologische' eigenschappen[211] – *politiek-maatschappelijke eigenschappen*, *religieuze eigenschappen*, *morele eigenschappen* – waarbij ik verwacht dat eigenschappen van de tweede categorie zeker tot de jaren tachtig structureel vaker voorkomen in Vlaamse recensies.

8.1 SYNCHRONE ANALYSE

Tabel 8.1 toont het gemiddelde aandeel van de eigenschapscodes in Nederlandse en Vlaamse recensies. Vrijwel alle verschillen zijn significant: alleen in het gebruik van de eigenschap *realiteitsgehalte/geloofwaardigheid* blijken Vlaamse en Nederlandse recensies niet te verschillen. Het grootste verschil vinden we bij de eigenschap die ook het grootste aandeel heeft in Vlaamse en Nederlandse recensies: *literaire/intellectuele eigenschappen*. Daarbij blijken Vlaamse critici zich vaker te buigen over 'de vraag of de besproken romans tot de literatuur behoorden, daarop aanspraak leken te willen maken en of ze literair-hoogstaand waren of niet', zoals Op de Beek deze eigenschap verder definieert.[212] Ook de eigenschap *originaliteit*, waarbij de oorspronkelijkheid van het werk onder de loep wordt gelegd, wordt in Vlaamse recensies vaker aan aspecten toegekend dan in Nederlandse recensies. In Nederlandse recensies, daarentegen, worden romans meer in algemene zin geëvalueerd, zoals aangeduid door het aandeel van *algemeen/overig*.

211 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 150.

212 Ibidem, 248.

Tabel 8.1: De resultaten van de eigenschapscodes over het gehele recensiecorpus. Gem. geeft de gemiddelde kans – uitgedrukt in percentages – voor elke eigenschapscode aan, SD de standaarddeviatie en de laatste drie kolommen zijn de statistieken van de Mann-Whitney toetsen met herkomst van de recensent als onafhankelijke variabele.

	Nederlands	Vlaams			
	Gem. (SD)	Gem. (SD)	U	p	r
Algemeen/overig	16.74 (9.92)	15.30 (8.47)	45445388.5	0.000	0.05
Continuïteit	1.61 (0.59)	1.67 (0.58)	39360382.0	0.000	-0.05
Didactische eigenschappen	0.94 (0.42)	0.97 (0.41)	40090167.0	0.000	-0.04
Efficiëntie	4.55 (3.93)	4.51 (3.32)	40560365.5	0.000	-0.03
Emotionaliteit	13.70 (8.18)	13.04 (7.13)	43514760.5	0.001	0.02
Helderheid	6.08 (4.04)	5.70 (3.57)	43979082.0	0.000	0.03
Humor	4.87 (6.18)	4.15 (5.16)	44370813.5	0.000	0.04
Intensiteit/levendigheid	9.34 (5.25)	9.75 (4.81)	39323598.0	0.000	-0.05
Literaire/intellectuele eigenschappen	17.06 (7.97)	19.38 (7.66)	34409189.0	0.000	-0.13
Morele eigenschappen	1.99 (1.55)	2.11 (1.50)	38642025.5	0.000	-0.06
Orginaliteit	5.30 (2.91)	5.70 (2.97)	37858839.0	0.000	-0.07
Politiek-maatschappelijke eigenschap.	2.05 (2.52)	2.30 (2.58)	37984374.0	0.000	-0.07
Realiteitsgehalte/geloofwaardigheid	8.95 (6.14)	8.61 (5.58)	42949177.0	0.058	0.01
Religieuze eigenschappen	0.82 (1.27)	0.76 (0.92)	39517129.5	0.000	-0.05
Samenhang	5.99 (4.13)	6.04 (3.75)	40898868.0	0.001	-0.02

Het verschil in *literair/intellectuele eigenschappen* en *originaliteit* sluit in zekere zin aan bij enkele ideeën die Hugo Brems – zij het met een focus op poëzie – heeft geformuleerd in 'De relatie tussen de Nederlandse en de Vlaamse literatuur'. De Vlaamse literatuur zou haar accenten leggen op 'decoratie en estheticisme', op experiment, terwijl er in Nederland, in de woorden van Luuk Gruwez in hetzelfde artikel, een 'fobie' zou zijn voor 'een literatuur die zich nadrukkelijk als literair afficheert', en functionaliteit geprefereerd wordt boven esthetiek.[213] Vanuit een dergelijke vermeende fobie zou het ook gemakkelijk te verklaren zijn waarom Nederlandse literatuurcritici minder aandacht besteden aan de vraag of de besproken romans al dan niet literair zijn.

Tegelijkertijd zijn Brems' bovenstaande observaties niet los te zien van de overkoepelende tegenstelling die hij in 'De relatie tussen de Nederlandse en de Vlaamse literatuur' formuleert, een tegenstelling die hij elf jaar later tot 'het bekende cliché' zou omdopen,[214] en die Ben van Humbeeck identificeert als 'een sjabloon die regelmatig blijkt terug te keren in de constructie van nationale identiteiten'.[215] De Vlaamse literatuur zou niet alleen meer gericht zijn op esthetiek en experiment, maar ook emotioneler en intuïtiever zijn, meer te maken hebben met 'expressie en belijdenis', terwijl in een Nederlandse context de accenten liggen op soberheid, organisatie, afstandelijkheid,

213 Brems, "De relatie tussen de Nederlandse en de Vlaamse literatuur. De teller en de noemer", 36-38.

214 Brems, "Het kritische discours over Vlaamse literatuur in Nederlandse kranten en weekbladen (1980-1994)", 107.

215 Ben Van Humbeeck, "'Het karakter van myten.'Literaire identiteitsvorming in de Vlaamse literaire kritiek (1980-1994)", *Neerlandistiek* 2007 (2007): 5.

constructie en helderheid.[216] Kortom: de tegenstelling tussen emotioneel en cerebraal. En dat (cliché)beeld wordt echter maar in beperkte mate bevestigd door de resultaten in Tabel 8.1. *Helderheid* komt inderdaad – al is het verschil relatief klein – significant vaker voor in Nederlandse recensies, en *intensiteit/levendigheid* vaker in Vlaamse recensies, maar daar blijf het ook bij. De andere affectief-georiënteerde eigenschappen – *emotionaliteit* en *humor* – komen juist vaker voor in Nederlandse recensies, en de eigenschappen die juist meer duiden op een aandacht voor organisatie, constructie en functionaliteit – respectievelijk *continuïteit*, *samenhang*, en *efficiëntie* – komen dan weer juist in Vlaams recensies vaker voor.

De ideologische eigenschappen, daarentegen, komen weliswaar minder vaak voor, maar laten wel een vrij eenduidig beeld zien. Zo komen *politiek-maatschappelijke eigenschappen*, zoals verwacht, inderdaad significant vaker in Vlaamse recensies. Sterker nog: op *literaire/intellectuele eigenschappen* en *originaliteit* na vinden we bij *politieke-maatschappelijke eigenschappen*, over het gehele corpus genomen, de grootste verschillen tussen Vlaamse en Nederlandse recensies – de aspectcodes incluis. Ook *didactische eigenschappen* en *morele eigenschappen* komen vaker voor in Vlaamse recensies. Het verschil in het aandeel van *religieuze eigenschappen* in Nederlandse en Vlaamse recensies is moeilijker te duiden. Het *gemiddelde* aandeel van *religieuze eigenschappen* is weliswaar hoger in Nederlandse recensies, maar er zijn – zo blijkt uit de Mann-Whitneytoets – systematisch meer Vlaamse dan Nederlandse recensies met een hoger aandeel in *religieuze eigenschappen*.[217] Dit laat vermoeden dat het relatief hoge Nederlandse gemiddelde aandeel in *religieuze eigenschappen* voornamelijk wordt bepaald door enkele uitschieters, die overwegend afkomstig lijken te zijn uit het *Reformatorisch Dagblad* (zie Bijlage A.1 voor het aandeel van de eigenschapscodes per krant). In Vlaamse recensies, daarentegen, lijkt er sprake van een minder sterke spreiding – daar heeft alleen *Het Handelsblad* een bovengemiddeld hoog aandeel in *religieuze eigenschappen*, maar deze krant is goed voor slechts één recensie in het recensiecorpus.

Daarmee komen dus alle ideologische eigenschappen vaker voor in Vlaamse recensies dan in Nederlandse recensies, waarbij we het grootste verschil in het aandeel van politiek-maatschappelijke evaluaties vinden. Een relevante vervolgvraag hierbij is in hoeverre er ook verschillen te vinden zijn in *de manier waarop* Nederlandse en Vlaamse critici de politieke eigenschappen van romans bespreken. Zo onderscheidt Op de Beek van *politiek-maatschappelijke eigenschappen* drie evaluatiecriteria: 'a) referentie aan de actualiteit, b) het engagement van de auteur (en de manier waarop daaraan vormgegeven

216 Brems, "De relatie tussen de Nederlandse en de Vlaamse literatuur. De teller en de noemer", 38.

217 Dat de effectgrootte een ander beeld laat zien dan het verschil tussen het gemiddelde aandeel in *religieuze eigenschappen* komt hoogstwaarschijnlijk doordat de Mann-Whitneytoets waarmee de effectgrootte is berekend niet werkt met gemiddeldes (in tegenstelling tot parametrische toetsen zoals de t-toets), maar met rangordes. Meer specifiek berekenen met een Mann-Whitneytoets of observaties van een groep significant vaker hoger in de rangorde staan dan die van een andere groep. Om die reden stelt Andy Field dan ook voor om in plaats van het gemiddelde het mediaan de rapporteren[218]. In ons geval is het mediaan van *religieuze eigenschappen* inderdaad ook hoger in Vlaamse recensies (0.46) dan in Nederlandse recensies (0.42). Om het onderzoek echter zo vergelijkbaar mogelijk te houden met dat van Linders en Op de Beek, heb ik er echter toch voor gekozen om, weliswaar met deze kanttekening, het gemiddelde te rapporteren.

wordt), c) het engageren van de lezer'.[219] Hoewel haar analyse van deze drie criteria hoofdzakelijk kwalitatief van aard is, kunnen we dus ook binnen politieke-maatschappelijke evaluaties variatie en verschil verwachten.

Om de politiek-maatschappelijke evaluaties in Vlaamse en Nederlandse recensies meer inhoudelijk te vergeleken, heb ik een vergelijkende *keyword*-analyse uitgevoerd op alle evaluatieve uitspraken waarvan de kans dat deze uitspraak met het label *politiek-maatschappelijke eigenschappen* geassocieerd wordt hoger is dan 0.5. Met een vergelijkende *keyword*-analyse kan geanalyseerd worden welke woorden significant vaker voorkomen in het ene deelcorpus (de Vlaamse recensies) dan in het andere deelcorpus (de Nederlandse recensies), en andersom. Daarbij kijk ik dus niet zozeer naar de woorden die in Nederlandse en Vlaamse politieke evaluaties het vaakst voorkomen (zie daarvoor Bijlage A.2), maar naar de woorden waarin Vlaamse en Nederlandse recensies het sterkst verschillen. Er zijn verschillen manieren om de woordfrequenties van twee corpora tegen elkaar af te zetten. In deze scriptie gebruik ik de chi-kwadraattoetsingsgrootheid (χ^2) als mijn centrale maatstaf. Tabel 8.2 toont de vijfentwintig meest significante woorden voor politieke evaluaties in Vlaamse (links) en Nederlandse (rechts) recensies.[220]

Het onderscheid dat Op de Beek maakt tussen politieke-maatschappelijke evaluaties die refereren aan de actualiteit en politieke-maatschappelijke evaluaties die zich richten op engagement, is allereerst duidelijk terug te zien in de tabel. In Vlaamse politieke evaluaties komt het woord 'engagement' significant vaker voor. Daarbij lijkt het vooral te gaan over het engagement van de auteur, en niet zozeer het engageren van de lezer. Zo schrijft Jeroen Versteele over Joseph Pearce: 'Elke gedachte, elke observatie getuigt van inzicht en engagement' (02-04-2008, *De Morgen*, Jeroen Versteele), en meent Willem M. Roggeman over Ben Borgart: 'Hij heeft duidelijk elk engagement uit zin boek willen weren en dit kan men in dit geval betreuren' (24-09-1979, *De Nieuwe Gazet*, Willem M. Roggeman).

In Nederlandse recensies, daarentegen, gaat het relatief minder over engagement, en komen woorden als 'actueel' en 'actualiteit' juist significant vaker voor. Vergeleken met evaluaties die het engagement van de auteur bespreken, is de relatie tot 'het politieke' in evaluaties die refereren aan de actualiteit betrekkelijk minder expliciet. Het actuele karakter van romans wordt – in Nederlandse recensies, althans – veelal relatief zijdelings benoemd, zoals 'haar thematiek blijft actueel' (21-02-2015, *Volkskrant*, Daniëlle Serdijn), 'een actueel boek' (19-09-2014, *Reformatorisch Dagblad*, Marie Verheij), of '*Meeuwen* is in ieder geval een actueel boek' (23-01-1976, *Algemeen/NRC Handelsblad*, Reinjan Mulders). Veel verder dan constateringen dat de thema's in bepaalde roman aansluiten bij recente nieuwswaardigheden gaan dergelijke evaluaties veelal niet – hoe een schrijver zich verhoudt tot die actualiteit, bijvoorbeeld, blijft in de meeste gevallen onbenoemd.

Naast 'engagement' komen ook woorden als 'Vlaams' en 'Vlaanderen' vaker in Vlaamse politieke evaluaties vaker voor. Dat lijkt in eerste instantie

219 Op de Beek, "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse", 156.

220 Enkele verdere technische specificaties van deze analyse: de deelcorpora zijn gelemmatiseerd en van POS-tags voorzien aan de hand van *Frog*, alle persoonsnamen zijn eruit gefilterd, en alle woorden met een frequentie lager dan vijf zijn niet in de figuur opgenomen.

Tabel 8.2: De 25 meest significante woorden voor politieke evaluaties in Vlaamse (links) en Nederlandse (rechts) recensies. Norm. freq geeft de per 10.000 woorden genormaliseerde frequentie in het desbetreffende subcorpus aan.

	Vlaamse politieke evaluaties			Nederlandse politieke evaluaties		
	Woord	Norm. freq.	χ^2	Woord	Norm. freq.	χ^2
1	toestand	15.58	23.94	lijken	20.77	11.18
2	korruptie	4.62	19.78	juist	13.71	9.92
3	precies	8.66	19.13	geven	37.42	9.58
4	hoofdpersonage	4.04	16.86	actualiteit	25.47	8.95
5	kultuur	4.62	15.99	actueel	40.95	8.08
6	repressie	4.62	15.99	komen	38.59	7.90
7	eeuwig	6.35	15.81	krijgen	24.49	7.85
8	aanklagen	4.04	13.21	bieden	8.23	7.35
9	overstijgen	5.19	13.12	ontwikkeling	11.75	7.10
10	engagement	42.12	12.60	hoofdpersoon	5.88	6.89
11	toe	14.43	12.22	aan_de_orde	5.88	6.89
12	vlaams	12.69	11.97	zaak	9.80	6.85
13	auteur	29.43	11.65	wel	51.92	6.75
14	hooggestemd	2.89	11.05	onderwerp	10.38	6.41
15	reflex	2.89	11.05	betrokkenheid	5.49	6.24
16	anekdotiek	2.89	11.05	stuk	9.99	5.86
17	nochtans	2.89	11.05	keuze	4.11	5.86
18	voedingsbodem	2.89	11.05	helder	3.92	5.52
19	situeren	5.19	11.01	verband	6.07	5.51
20	typisch	6.92	10.93	jaar	28.02	5.46
21	bespiegeling	4.62	10.67	verandering	4.90	5.27
22	evenwel	3.46	10.48	waaraan	3.72	5.18
23	o.m.	3.46	10.48	opzicht	5.68	4.90
24	wegen	3.46	10.48	moment	7.44	4.76
25	Vlaanderen	6.92	9.61	waarheid	6.47	4.65

wellicht weinig betekenisvol – Vlaamse critici schrijven nu eenmaal meer over Vlaamse auteurs dan Nederlandse critici.[221] Maar waar Nederlandse critici evenzeer meer over Nederlandse auteurs schrijven, zijn er geen significante verschillen tussen Vlaamse en Nederlandse critici in het gebruik van woorden zoals 'Nederlands' of 'Nederland' in politieke evaluaties.[222] Politieke evaluaties worden in Vlaamse recensies dus relatief vaak ook met Vlaamse kwesties verbonden, terwijl de Nederlandse context in Nederlandse recensies minder zwaar weegt, of in ieder geval zo vanzelf spreekt dat hij niet wordt benoemd.

Er zijn verschillende manieren waarop de Vlaamse context gestalte krijgt in politieke evaluaties. Allereerst zien we enkele referenties aan het Vlaamsnationalisme en Vlaamsnationalistische bewegingen – Vlaams Blok, Vlaams Belang, de Vlaamse beweging – waarbij de nadruk ligt op de vraag in hoeverre de auteur geslaagd (hetgeen vaak neerkomt op gelaagd) uiting heeft gegeven aan deze bewegingen. Zo zou Dimitri Verhulst op een clichématige manier 'de agressieve Vlaams-nationalist' schetsen (23-09-2011, *De Standaard*, Mark Cloostermans), terwijl *Los* van Tom Naegels wordt geprezen om zijn

221 Zie ook Van Humbeeck en Van Renssen, "Op voet van gelijkheid? De Noord-Zuidkwestie in de Nederlandse en de Vlaamse literaire kritiek, 1980-1995".

222 *Nederland*: $\chi^2 = 3.28, p = 0.07$; *nederlands*: $\chi^2 = 0.98, p = 0.32$

nauwgezette (psychologische) analyse van het Vlaams Belang: '*Los* is geen eenduidig boek over het Vlaams Belang maar over wat daaronder ligt: de psychologie die ons handelen drijft, en die we in politieke termen uitdrukken om ze iets te verlenen waar ze eigenlijk geen recht op heeft: een schijn van objectiviteit' (02-04-2005, *De Tijd*, Jeroen Overstijns). Veel vaker zien we echter referenties aan de manier waarop de auteur uiting heeft gegeven aan een specifiek Vlaams milieu, zoals in Borré's bespreking van *Het verdriet van België* aan het begin van dit hoofdstuk. Interessant aan dergelijke referenties is ook dat ze niet zelden gepaard gaan met het idee dat het schetsen van een repressief milieu karakteristiek is voor de Vlaamse literatuur. Zo is het koppelen van 'triviale familiale toestanden' aan 'geschiedenis en/of politiek' volgens Mark Cloostermans 'typisch voor de gemiddelde Vlaamse roman' (09-09-2016, *De Standaard*, Mark Cloostermans), spreekt Jos Borré van 'ons typische miserabilisme' dat in *Eens kind , altijd kind*, al wroetende in 'een aloude Vlaamse traditie', naar boven komt (07-07-2004, De Morgen, Jos Borré), en schrijft Jeroen Overstijns:

> Neem bijvoorbeeld de stempel die het rooms-katholicisme op het Vlaamse intellectuele leven heeft gedrukt. [...] Morele en psychologische frustratie, het slachtofferschap, het eeuwige verhaal van schuld en boete, ze zetten nog altijd in niet geringe mate de lijnen van de Vlaamse literatuur uit. Dat is niet anders in *Het leeuwerikshuis* van Annie van Keymeulen. (26-03-1998, *De Standaard*, Jeroen Overstijns)

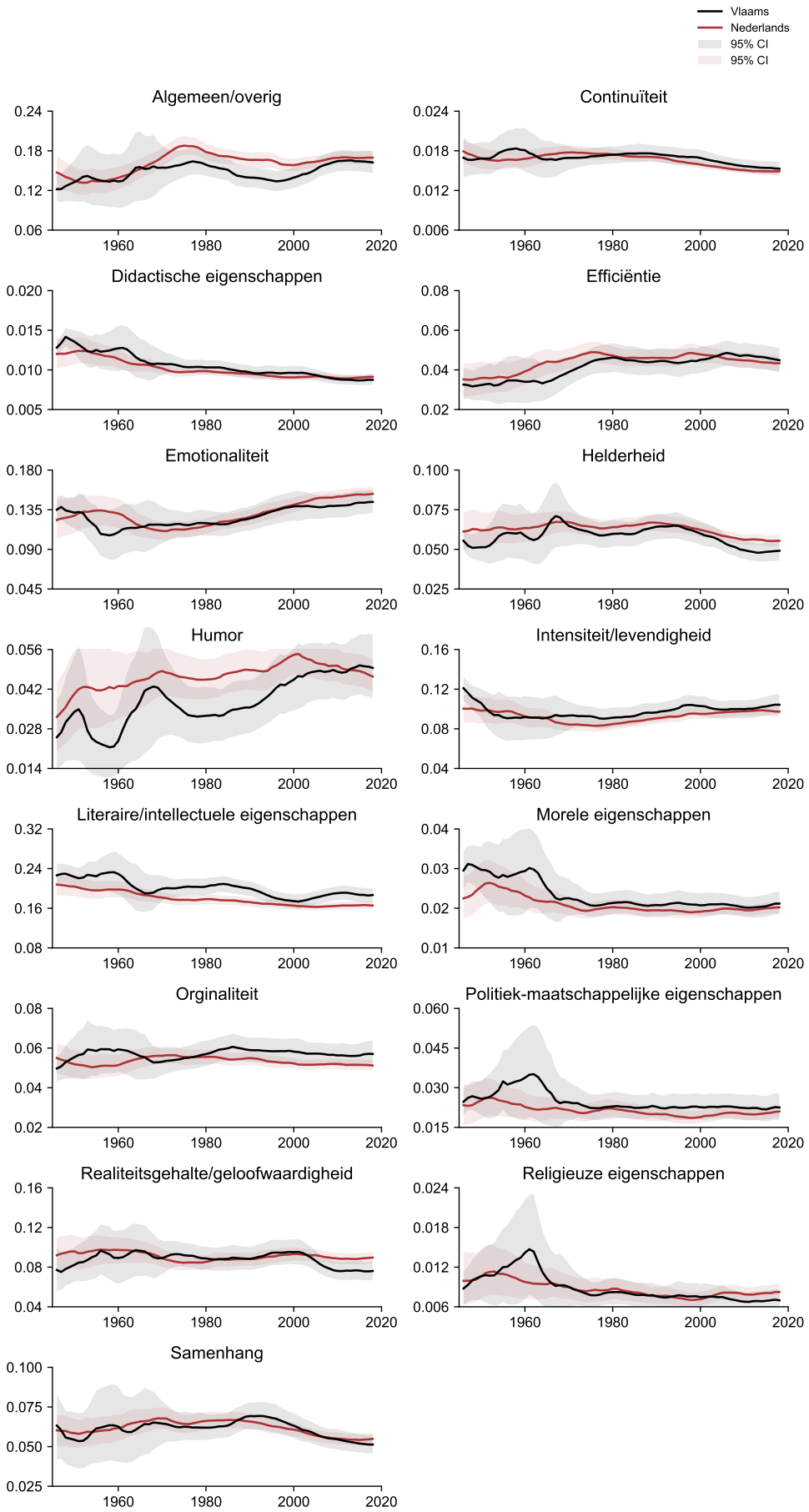
Waarbij – zeker in het geval van Overstijns – het aansluiten bij een dergelijke Vlaams-literaire traditie niet noodzakelijk positief wordt gewaardeerd, maar eigenlijk eerder wordt beschouwd als een gebrek aan originaliteit.

Kijkend naar de rol die 'het politieke' speelt in Vlaamse en Nederlandse literatuurkritiek, kunnen we dus niet alleen bevestigen dat Vlaamse critici inderdaad meer aandacht hebben voor de politiek-maatschappelijke eigenschappen van proza, maar ook dat de manier waarop Nederlandse en Vlaamse critici de politiek-maatschappelijke eigenschappen bespreken verschilt. Waar Nederlandse critici meer naar de actualiteit refereren, spreken Vlaamse critici juist meer over engagement, en worden politieke opvattingen in literaire werken verbonden aan een specifiek Vlaams milieu.

8.2 DIACHRONE ANALYSE

Figuur 8.1 toont de proportionele aandacht voor de zestien eigenschapscodes in Vlaamse en Nederlands recensies door de tijd heen. In tegenstelling tot bij de aspectcodes – waar het aandeel van bijvoorbeeld *plot* bijna verdriedubbelde – blijken bij de eigenschappen aanmerkelijk minder grote verschillen tussen jaren te zijn. Alleen het aandeel van *humor* is tussen 1946 en 2000 in Nederlandse én Vlaamse recensies verdubbeld.

Hoewel de concrete toename dus beperkt is, ontwikkelen de werkimmanente eigenschappen in Nederlandse recensies zich in grote lijnen op dezelfde manier. Het aandeel van zowel *samenhang*, *continuïteit*, als dat van



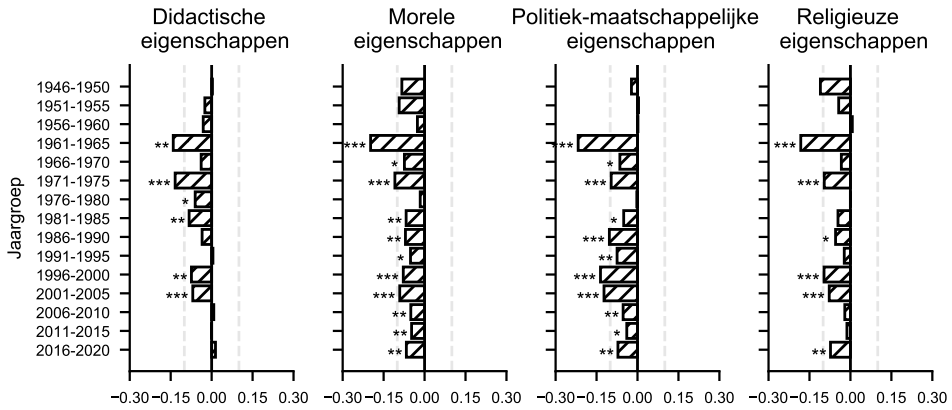
Figuur 8.1: De proportionele aandacht voor de eigenschapscodes door de tijd heen. Zie ook Figuur 7.2 voor een toelichting op de vorm van deze figuur.

helderheid neemt vanaf 1946 aanvankelijk (beperkt) toe, maar vanaf het einde van de jaren tachtig, begin jaren negentig, weer af. Zo neemt het aandeel van *samenhang* tussen 1950 en 1970 met ongeveer 16% toe, het aandeel van *helderheid* met ongeveer 8%, en dat van *continuïteit* met (slechts) 5%. Tussen 1985 en 2018 neemt het aandeel van *samenhang* dan weer met 21% af, van *helderheid* met 18%, en van *continuïteit* met 14%. Ook bij de eigenschappen *originaliteit* en *efficiëntie* is er in Nederlandse recensies er sprake van een soortgelijk patroon. In Vlaamse recensies verloopt de toename van deze werkimmanente eigenschappen grilliger, maar zien we wel pas later – midden jaren negentig – een definitieve afname in het aandeel van de werkimmanente eigenschappen. Daarbij neemt het aandeel *samenhang* in Vlaamse recensies vooral tussen 1980 en 1990 toe, met ongeveer 12%. In deze periode zien we in Vlaamse recensies eveneens een lichte toename in het aandeel van de eigenschap *helderheid*.

Bij de eigenschappen *emotionaliteit* en *realiteitsgehalte/geloofwaardigheid*, die niet zozeer betrekking hebben op het werk zelf, is er in Nederlandse recensies sprake van een omgekeerd patroon aan dat van de werkimmanente eigenschappen. Het aandeel van *emotionaliteit* en *realiteitsgehalte/geloofwaardigheid* neemt vanaf 1955 tot midden jaren zeventig juist af, om daarna weer toe te nemen. Zeker voor *emotionaliteit* is deze toename relatief sterk: tussen 1970 en 2018 is het aandeel hiervan met 38% toegenomen. In Vlaamse recensies neemt de aandacht voor *emotionaliteit* daarentegen al vanaf 1955 gestaag toe, terwijl de aandacht voor *realiteitsgehalte/geloofwaardigheid* vanaf 2000 juist relatief sterk afneemt.

Het aandeel van *algemeen/overig* neemt lichtelijk toe, in Nederlandse recensies het sterkst vanaf 1950 tot 1975, in Vlaamse recensies vanaf 1995 tot 2013, al blijven algemene evaluaties structureel vaker voorkomen in Nederlandse recensies. Voor *literaire/intellectuele eigenschappen* geldt het omgekeerde: daarvan neemt het aandeel in Nederlandse en Vlaamse recensies lichtelijk af, maar blijven Vlaamse recensies door de jaren heen consequent meer aandacht besteden aan de literariteit van de besproken roman.

Waar de verschillen tussen Vlaamse en Nederlandse recensies bij eigenschappen zoals *efficiënte*, *emotionaliteit* of *continuïteit* uiteindelijk behoorlijk fluctueren, zien we bij de ideologische eigenschappen (wederom) de meest structurele patronen. Zowel *morele eigenschappen*, *didactische eigenschappen*, als *politiek-maatschappelijke eigenschappen* blijken over bijna de gehele periode vaker in Vlaamse recensies voor te komen dan in Nederlandse recensies. Alleen het gemiddelde aandeel van *religieuze eigenschappen* is in recentere Nederlandse recensies hoger. Daarbij neemt het aandeel van deze eigenschappen in Nederlandse recensies sinds 1946 over het algemeen af, is er vooral bij *politiek-maatschappelijke eigenschappen* rond 1973 en 1998 sprake van twee momenten van lichte stijging. In Vlaamse recensies is er bij *morele eigenschappen*, *religieuze-eigenschappen* en *politiek-maatschappelijke eigenschappen* daarentegen tot midden jaren zestig eerst nog sprake van een toename, en zien we het aandeel van deze eigenschappen daarna relatief snel afnemen, waarna het bij *morele eigenschappen* en *politiek-maatschappelijke eigenschappen* vrij constant blijft, en bij *religieuze eigenschappen* nog lichtelijk afneemt. Rond de eerste helft van de jaren zestig kunnen we dan ook de grootste verschillen



Figuur 8.2: De vijfjaarlijkse effectgroottes (r) van de ideologische eigenschappen. Een positieve waarde geeft aan dat een eigenschap in Nederlandse recensies een hoger aandeel heeft; een negatieve waarde geeft aan dat een eigenschap in Vlaamse recensies een hoger aandeel heeft. *** $p < 0.001$; ** $p < .01$; * $p < .05$

tussen Vlaamse en Nederlandse recensies in het aandeel van de ideologische eigenschappen waarnemen.

Wanneer we het recensiecorpus, op dezelfde manier als in Hoofdstuk 7.2, opdelen in periodes van vijf jaar, en voor elke periode de verschillen tussen Nederlandse en Vlaamse recensies bij de ideologische eigenschappen apart toetsen (zie Figuur 8.2), zien we eens te meer hoe structureel de verschillen op deze eigenschappen zich voordoen. In geen enkele periode komen evaluaties die betrekking hebben op *didactische eigenschappen*, *morele eigenschappen*, *politiek-maatschappelijke eigenschappen* of *religieuze eigenschappen* in Nederlandse recensies significant vaker voor dan in Vlaamse recensies. Andersom zien we evaluaties die betrekking hebben op *didactische eigenschappen* en *religieuze eigenschappen* in zes van de vijftien periodes significant vaker in Vlaamse recensies voorkomen, en evaluaties die betrekking hebben op *morele eigenschappen* en *politiek-maatschappelijke eigenschappen* zelfs in elf van de vijftien periodes. Voor alle vier eigenschappen is de effectgrootte het veruit het hoogst in de periode tussen 1961 en 1965. Bij *politiek-maatschappelijke eigenschappen* is er in deze periode sprake van een effectgrootte van -0.22, bij *morele eigenschappen* van -0.20, bij religieuze eigenschappen van -0.18, en bij *didactische eigenschappen* van -0.14. Ook in de periode 1971-1975 treffen we bij alle vier eigenschappen een relatief hoge effectgrootte aan, terwijl in 1976-1980, afgezien *didactische eigenschappen* geen verschillen tussen Vlaamse en Nederlandse recensies zichtbaar zijn. Omdat er echter in de periodes daarna wel weer sprake is van significante verschillen – zeker bij *politiek-maatschappelijke eigenschappen* in 1996-2000 en 2001-2005 – kunnen we in deze periode niet eenduidig een effect van de Vlaamse ontzuiling waarnemen. Eerder toont de figuur aan dat ideologische eigenschappen vrij consequent vaker in Vlaamse recensies voorkomen.

8.3 VERSCHILLEN TUSSEN NEDERLANDSE EN VLAAMSE SCHRIJVERS

Waar de diachrone analyse bij de eigenschappen minder duidelijke verschillen tussen jaren laat zien dan bij de aspecten, zijn er bij de dubbel-comparatieve analyse, waarbij we ook de herkomst van auteur in de analyse verdisconteren,[223] juist meer (significante) verschillen te waar te nemen (zie Tabel 8.3).[224]

Zo zagen we bij de aspecten in Nederlandse recensies geen verschillen in de manier waarop Vlaamse en Nederlandse auteurs besproken worden. Dat is bij de eigenschappen niet het geval. Naast *algemeen/overig* zien we bij alle vier ideologische eigenschappen – *didactische eigenschappen*, *morele eigenschappen*, *politiek-maatschappelijke eigenschappen* én *religieuze eigenschappen* – een significant verschil in de manier waarop Nederlandse en Vlaamse auteurs in Nederlandse recensies besproken worden. Daarbij komen de ideologische eigenschappen – al zijn de verschillen relatief beperkt – alle vier vaker voor in besprekingen van Vlaamse auteurs.

Tabel 8.3: Effectgroottes (r) van de paarsgewijze vergelijkingen waarbij groepen niet zijn vergeleken aan de hand van de herkomst van de recensent, maar ook de herkomst van de auteur.

Label	Groep	Gem.	SD	1	2	3	4
Algemeen/overig	1. Ned. criticus, Ned. auteur	16.89	10.00	-			
	2. Ned. criticus, Vl. aut.	15.50	9.20	-0.04	-		
	3. Vl. criticus, Ned. auteur	15.14	8.01	-0.05	ns	-	
	4. Vl. criticus, Vl. auteur	15.46	8.88	-0.05	ns	ns	-
Continuïteit	1. Ned. criticus, Ned. auteur	1.61	0.59	-			
	2. Ned. criticus, Vl. aut.	1.61	0.58	ns	-		
	3. Vl. criticus, Ned. auteur	1.62	0.57	ns	ns	-	
	4. Vl. criticus, Vl. auteur	1.71	0.60	0.07	0.09	0.08	-
Didactische eigenschappen	1. Ned. criticus, Ned. auteur	0.94	0.41	-			
	2. Ned. criticus, Vl. aut.	0.97	0.42	0.03	-		
	3. Vl. criticus, Ned. auteur	0.94	0.40	ns	ns	-	
	4. Vl. criticus, Vl. auteur	0.99	0.42	0.05	ns	0.07	-
Efficiëntie	1. Ned. criticus, Ned. auteur	4.54	3.95	-			
	2. Ned. criticus, Vl. aut.	4.61	3.76	ns	-		
	3. Vl. criticus, Ned. auteur	4.35	3.19	ns	ns	-	
	4. Vl. criticus, Vl. auteur	4.67	3.44	0.04	0.04	0.05	-
Emotionaliteit	1. Ned. criticus, Ned. auteur	13.67	8.20	-			
	2. Ned. criticus, Vl. aut.	13.97	7.98	ns	-		
	3. Vl. criticus, Ned. auteur	13.60	7.49	ns	ns	-	
	4. Vl. criticus, Vl. auteur	12.50	6.73	-0.04	-0.08	-0.07	-

Vervolg tabel op de volgende pagina

223 Zie ook Hoofdstuk 7.3 voor een verdere specificatie.

224 Op *realiteitsgehalte/geloofwaardigheid* na is er voor elk label sprake van een hoofdeffect. *Algemeen/overig*: H(4)=100.45, p<0.001; *Continuïteit*: H(4)=80.3, p<0.001; *Didactische eigenschappen*: H(4)=62.62, p<0.001; *Efficiëntie*: H(4)=29.73, p<0.001; *Emotionaliteit*: H(4)=36.19, p<0.001; *Helderheid*: H(4)=22.42, p<0.001; *Humor*: H(4)=39.08, p<0.001; *Intensiteit/levendigheid*: H(4)=65.24, p<0.001; *Literaire/intellectuele eigenschappen*: H(4)=401.8, p<0.001; *Morele eigenschappen*: H(4)=128.28, p<0.001; *Orginaliteit*: H(4)=120.57, p<0.001; *Politiek-maatschappelijke eigenschappen*: H(4)=170.39, p<0.001; *Realiteitsgehalte/geloofwaardigheid*: H(4)=4.12, p=0.249; *Religieuze eigenschappen*: H(4)=77.88, p<0.001; *Samenhang*: H(4)=23.47, p<0.001

Tabel 8.3: Effectgroottes (r) van de paarsgewijze vergelijkingen waarbij groepen niet zijn vergeleken aan de hand van de herkomst van de recensent, maar ook de herkomst van de auteur.

Label	Groep	Gem.	SD	1	2	3	4
Helderheid	1. Ned. criticus, Ned. auteur	6.10	4.06	-			
	2. Ned. criticus, Vl. aut.	5.93	3.86	ns	-		
	3. Vl. criticus, Ned. auteur	5.78	3.66	-0.02	ns	-	
	4. Vl. criticus, Vl. auteur	5.62	3.48	-0.03	ns	ns	-
Humor	1. Ned. criticus, Ned. auteur	4.85	6.19	-			
	2. Ned. criticus, Vl. aut.	5.03	6.14	ns	-		
	3. Vl. criticus, Ned. auteur	4.44	5.44	ns	ns	-	
	4. Vl. criticus, Vl. auteur	3.86	4.86	-0.04	-0.07	-0.04	-
Intensiteit/levendigheid	1. Ned. criticus, Ned. auteur	9.32	5.26	-			
	2. Ned. criticus, Vl. aut.	9.52	5.18	ns	-		
	3. Vl. criticus, Ned. auteur	10.01	5.02	0.06	0.06	-	
	4. Vl. criticus, Vl. auteur	9.49	4.59	0.03	ns	-0.04	-
Literaire/intellectuele eigenschappen	1. Ned. criticus, Ned. auteur	17.00	7.94	-			
	2. Ned. criticus, Vl. aut.	17.51	8.22	ns	-		
	3. Vl. criticus, Ned. auteur	19.02	7.71	0.09	0.11	-	
	4. Vl. criticus, Vl. auteur	19.74	7.60	0.13	0.16	0.05	-
Morele eigenschappen	1. Ned. criticus, Ned. auteur	1.97	1.53	-			
	2. Ned. criticus, Vl. aut.	2.17	1.74	0.04	-		
	3. Vl. criticus, Ned. auteur	2.02	1.44	0.03	ns	-	
	4. Vl. criticus, Vl. auteur	2.20	1.55	0.08	0.05	0.07	-
Originaliteit	1. Ned. criticus, Ned. auteur	5.31	2.91	-			
	2. Ned. criticus, Vl. aut.	5.25	2.85	ns	-		
	3. Vl. criticus, Ned. auteur	5.69	3.07	0.06	0.09	-	
	4. Vl. criticus, Vl. auteur	5.71	2.87	0.06	0.1	ns	-
Politiek-maatschappelijke eigenschappen	1. Ned. criticus, Ned. auteur	2.03	2.53	-			
	2. Ned. criticus, Vl. aut.	2.20	2.46	0.03	-		
	3. Vl. criticus, Ned. auteur	2.10	2.43	0.03	ns	-	
	4. Vl. criticus, Vl. auteur	2.49	2.70	0.09	0.08	0.09	-
Realiteitsgehalte/geloof-waardigheid	1. Ned. criticus, Ned. auteur	8.96	6.14	-			
	2. Ned. criticus, Vl. aut.	8.90	6.12	ns	-		
	3. Vl. criticus, Ned. auteur	8.68	5.68	ns	ns	-	
	4. Vl. criticus, Vl. auteur	8.55	5.48	ns	ns	ns	-
Religieuze eigenschappen	1. Ned. criticus, Ned. auteur	0.81	1.28	-			
	2. Ned. criticus, Vl. aut.	0.84	1.19	0.03	-		
	3. Vl. criticus, Ned. auteur	0.72	0.92	0.02	ns	-	
	4. Vl. criticus, Vl. auteur	0.80	0.91	0.06	ns	0.06	-
Samenhang	1. Ned. criticus, Ned. auteur	5.99	4.15	-			
	2. Ned. criticus, Vl. aut.	6.00	3.98	ns	-		
	3. Vl. criticus, Ned. auteur	5.89	3.73	ns	ns	-	
	4. Vl. criticus, Vl. auteur	6.20	3.76	0.04	0.05	0.05	-

Ook in Vlaamse recensies is er bij besprekingen van Vlaamse auteurs meer aandacht voor de ideologische eigenschappen, en in het bijzonder voor *politiek-maatschappelijke eigenschappen*. Sterker nog: de verschillen tussen Vlaamse recensies zijn groter dan die tussen Vlaamse en Nederlandse recensies.

Dat zowel Nederlandse als Vlaamse critici bij Vlaamse auteurs meer aandacht voor de ideologische eigenschappen hebben, suggereert dat deze eigenschappen ook meer relevant zijn bij Vlaamse auteurs, dat Vlaamse romans – met andere woorden – in zekere zin ook meer religieus, politiek,

didactisch dan wel moreel van aard zijn. Vanuit dat perspectief zou mijn analyse niet zozeer verschillende poëticale opvattingen weerspiegelen, maar vooral een verschillend aanbod – Vlaamse critici schrijven nu eenmaal vaker over Vlaamse auteurs, en Vlaamse auteurs schrijven nu eenmaal vaker over politiek.[225] Toch is dat perspectief niet volledig juist. Zeker bij *politiek-maatschappelijke eigenschappen* en *morele eigenschappen* verklaart de herkomst van de auteur allerminst alle verschillen tussen Vlaamse en Nederlandse critici. Zo worden Vlaamse auteurs nog steeds veel politieker in Vlaamse recensies dan in Nederlandse recensies beoordeeld, en is het aandeel van *politiek-maatschappelijke eigenschappen* bij besprekingen van Nederlandse auteurs eveneens (beperkt, maar significant) hoger in Vlaamse recensies.

Naast de ideologische eigenschappen is er in Vlaamse recensies bij *continuïteit*, *efficiëntie*, *samenhang*, *literaire/intellectuele eigenschappen*, *emotionaliteit*, *humor* en *intensiteit/levendigheid* eveneens sprake van een significant verschil tussen Nederlandse en Vlaamse auteurs. Verrassend, gegeven de verschillen bij de ideologische eigenschappen, is dat de affectieve eigenschappen – *emotionaliteit*, *intensiteit/levendigheid*, en *humor* – juist vaker aan bod komen bij besprekingen van Nederlandse auteurs, terwijl de werkimmanente eigenschappen – *continuïteit* en *samenhang* – juist vaker voorkomen bij besprekingen van Vlaamse auteurs. Vlaamse recensenten beoordelen Nederlandse auteurs dus niet eenduidig vanuit meer een autonomistisch poëticaal kader dan Vlaamse auteurs. Bij Vlaamse auteurs is er meer aandacht voor ideologie, bij Nederlandse auteurs ligt de nadruk eerder op expressie. Wel kunnen we, de resultaten van de dubbel-comparatieve analyse van de aspecten eveneens in acht genomen, concluderen dat de herkomst van de auteur voor Vlaamse critici zwaarder meeweegt dan bij Nederlandse critici.

225 Natuurlijk ligt de zaak complexer dan ik hier – bewust hyperbolisch – opschrijf. Literatuurcritici zijn uiteraard, het verhaal zal bekend zijn, geen neutrale doorgeefluiken van "het" aanbod, maar hebben een poortwachtersfunctie en kiezen mede vanuit hun poëticale opvattingen boeken uit om te recenseren.

Deel IV

CONCLUSIE

Hoofdstuk 9

CONCLUSIE

'Een wetenschappelijk verantwoorde geschiedschrijving dient een geheel te beschrijven, dient een globaal beeld op te leveren. De beschrijving van de afzonderlijke bomen of boompartijen moet het zicht op het bos niet verhinderen'.[226] Aldus Anne Marie Musschoot nadat ze Ton Anbeek, die in zijn literatuurgeschiedenis uit 1990 de 'radicale beslissing' maakt de Nederlandse en Vlaamse literatuur als 'aparte grootheden' te beschouwen,[227] verwijt geen oog te hebben voor (literair-historische) voorbeelden die niet duiden op verschil, maar op – ik citeerde het al eerder – 'verwevenheid, samenhang en eenheid'.[228] Niet de bomen, maar het bos – als we niet beter wisten, had dit net zo goed een argument voor 'distant reading' *avant la lettre* kunnen zijn. Maar voor een kwantitatieve of computergestuurde benadering kiest Musschoot uiteraard niet. Musschoot haalt tegenvoorbeelden aan, om daarmee te beargumenteren dat een 'literaire geschiedenis van de Nederlanden' onvermijdelijk rekening moet houden 'met de Vlaamse component'.[229] Of, om in haar metafoor te blijven: ze wijst enkele andere bomen aan om daarmee te beargumenteren dat het bos er toch wel anders uit zou moeten zien dan Anbeek beweert.

Tegenstanders van computationele benaderingen in de letterkunde wijzen er maar al te vaak op dat het lezen van teksten met de computer op een problematische manier reductionistisch is. 'CLS [computational literary studies]', schrijft Nan Z. Da bijvoorbeeld stellig, 'has no ability to capture literature's complexity'.[230] Maar ook kwalitatieve lezingen, zo kunnen we uit de gedachtewisseling tussen Anbeek en Musschoot concluderen, zijn allerminst gevrijwaard van reductionistische problematiek. Een cruciaal element van close reading is het *selecteren* van de juiste voorbeelden, van voorbeelden die, aldus Paul Fleming, 'larger structures and contexts' illustreren, waarbij de vraag in hoeverre het geheel gereduceerd kan worden tot deze voorbeelden altijd onderwerp van interpretatieve strijd blijft – zo ook bij Anbeek en

226 Musschoot, "Anbeeks keuze: enkele kanttekeningen", 235.
227 Anbeek, *Geschiedenis van de Nederlandse literatuur tussen 1885 en 1985*, 16.
228 Musschoot, "Anbeeks keuze: enkele kanttekeningen", 235.
229 Ibidem, 236.
230 Da, "The Computational Case against Computational Literary Studies", 634.

Musschoot.[231] En dat bedoel ik geenszins als een epistemologische *tu quoque*, maar als een duiding van verschil, van toegevoegde waarde. Ook kwalitatieve analyses zijn – tot op zekere hoogte – reductionistisch. Maar waar distant reading het deel – individuele romans, recensies, evaluatieve uitspraken – reduceert tot telbare eenheden, reduceert close reading het geheel tot analyseerbare voorbeelden.

In deze scriptie koos ik voor een computationele aanpak om moderne Nederlandse en Vlaamse literatuurkritiek als geheel te kunnen vergelijken. Anders dan het gros van het letterkundig onderzoek dat gebruik maakt van digitale technieken koos ik niet voor reeds bestaande computationele analysetechnieken, voor de analyse van *n-grams*, *topics*, *sentiment* of *named entities*, maar stelde ik mij methodologisch ten doel het categoriseringsmodel voor literaire recensies dat is ontwikkeld door Yvette Linders en Esther op de Beek automatisch te reproduceren. Ik vergeleek daarvoor, in het tweede deel van deze scriptie, verschillende manieren om zinnen numeriek te representeren en modellen om op basis van deze representaties zinnen te categoriseren, variërend van heel simpel tot vrij complex. Het Nederlandstalige BERT-model, het meest complexe model, bleek zowel voor het classificeren van evaluatieve uitspraken, als voor het ranken van aspect- en eigenschapscodes de meest accurate resultaten op te leveren.

Het BERT-model is een contextueel taalmodel: de numerieke representatie van een woord is niet gefixeerd, maar afhankelijk van de context waarin dit woord voorkomt. En hoewel ik heb laten zien dat vectorrepresentaties die getraind zijn op literaire recensies in staat zijn subtiele literair-historische verhoudingen te reproduceren, is deze context natuurlijk ook niet onbegrensd. Alle modellen hadden – in mijn configuratie, althans – bij het voorspellen van de labels van een zin ook alleen deze zin tot hun beschikking, en niet de rest van de recensie. Een menselijke codeur heeft een dergelijke context wel; die kan ervoor kiezen een recensie eerst van voor tot achter door te lezen, en daarna pas een beslissing te maken. Daarmee, zou je kunnen beargumenteren, is zelfs de inzet van het meest complexe taalmodel, bestaande uit miljoenen parameters, een (al dan niet drastische) reductie van de complexiteit van evaluatieve uitspraken. En dat geldt eigenlijk evenzeer voor het onderbrengen van evaluatieve uitspraken in twaalf aspecten en zestien eigenschappen *an sich*, hetgeen niet altijd recht doet aan de grote variëteit aan literair-kritische uitspraken.[232]

Die dubbele reductie wil ik dan ook zeker niet ontkennen. In tegendeel: ik acht haar eerder noodzakelijk. Noodzakelijk, niet alleen om telbare eenheden in recensies te kunnen onderscheiden, maar ook om dat op een efficiënte wijze in meer dan twintigduizend recensies te kunnen doen. Maar waar het reductionisme van de methodiek die ik in deze scriptie heb gebruikt zich onderscheidt van veel andere computationele technieken, is dat het bij uitstek een vorm van gevalideerd reductionisme is. Voor elk model heb ik immers

231 Paul Fleming, "Tragedy, for Example: Distant Reading and Exemplary Reading (Moretti)", *New Literary History* 48, nummer 3 (2017): 437.

232 Vandaar ook dat ruim een kwart van evaluatieve uitspraken in een restcategorie terechtkomt, zoals Sven Vitse in zijn recensie van Op de Beeks proefschrift schrijft. Zie Vitse, "Esther op de Beek, Een literair fenomeen van de eerste orde. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: Een kwantitatieve en kwalitatieve analyse.", 149.

geëvalueerd in hoeverre de voorspellingen overeenkomen met de codes die door menselijke codeurs zijn toegekend. Hierdoor weten we vrij precies in hoeverre wat we meten overeenkomt met wat we willen meten. Dat is, zoals ik het eerste deel van deze scriptie betoogde, anders bij ongecontroleerde (*unsupervised*) zelflerende algoritmes, zoals *topic modeling*, waar een toename van thema rond politieke woorden nog niet noodzakelijk duidt op een toename van politieke evaluaties.

En met een dergelijke gevalideerde reductie verliezen we mijns inziens niet het zicht op complexiteit, maar kunnen we er juist beter grip op krijgen. Dat is dan niet zozeer de complexiteit van het deel, de complexiteit van 'irony, ambiguity and paradox' of gelijkaardige opsommingen van "typisch" literaire stijlmiddelen,[233] maar de complexiteit van het geheel; een eerder numerieke complexiteit. Deze scriptie moet dan ook allerminst gelezen worden als een poging eindelijk empirisch aan te tonen dat de Nederlandse en de Vlaamse literatuur(kritiek) één danwel twee vormen. Als ik al, op basis van de resultaten in het vierde deel van deze scriptie, zo'n antwoord zou willen geven, dan komt dit nog het meest in de buurt van het 'allicht eenvoudig[e]' antwoord dat Geert Buelens in zijn pleidooi voor een vergelijkende neerlandistiek heeft geformuleerd: 'het hangt er maar vanaf hoe en wanneer' je ernaar kijkt.[234] Want met achtentwintig variabelen (synchroon en diachroon vergeleken) is de verhouding tussen Nederlandse en Vlaamse literatuurkritiek geen kwestie van bij elkaar dan wel apart, geen binair of/of, maar zien we juist continue, soms ook tegenstrijdige, pendelbewegingen tussen convergentie en divergentie, tussen verschil en overeenkomst.

Eerder dan aan te tonen *dat* de Vlaamse en Nederlandse literatuurkritiek al dan niet bij elkaar horen, ging het mij in deze scriptie om het beschrijven, analyseren en – voor zover dat mogelijk bleek – duiden van die verschillen en overeenkomsten. Daarbij vertrok ik vanuit de hypothese dat de nauwe verwevenheid van de Vlaamse Beweging met de Vlaamse literatuur ervoor zorgde dat Vlaamse literatuurcritici literaire werken vanuit een minder autonoom poëticaal kader beoordelen dan Nederlandse critici. Eveneens verwachtte ik dat, vanwege de latere ontzuiling in Vlaanderen en de invloed van *Merlyn* in Nederland in de jaren zestig, buitenliteraire factoren langer hun stempel zouden drukken op de oordelen van Vlaamse recensies, en de aandacht voor werkimmanente aspecten ook pas later toeneemt. Om deze verwachtingen te toetsen, classificeerde ik evaluatieve uitspraken in een corpus van 22.361 Nederlandse en Vlaamse recensies van Nederlandstalig proza op basis van het categoriseringsmodel van Linders en Op de Beek.

Ideologische eigenschappen bleken, allereerst, inderdaad structureel vaker aan romans te worden toegekend in Vlaamse recensies dan in Nederlandse recensies. Over bijna de gehele periode komen evaluaties die betrekking hebben op *didactische eigenschappen, morele eigenschappen, politiek-maatschappelijke eigenschappen* en *religieuze eigenschappen* significant vaker voor in Vlaamse recensies. Daarbij is het verschil tussen Nederlandse en Vlaamse recensies in het aandeel van deze ideologische eigenschappen het grootst tussen 1961 en 1965. Van alle ideologische eigenschappen verschillen Nederlandse en Vlaamse recensies het sterkst in het gebruik van *politiek-maatschappelijke*

233 Elizabeth Freund, *Return Of Reader* (Routledge, 2013), 41.
234 Buelens, "Naar een Vergelijkende Neerlandistiek".

eigenschappen. Een vergelijkende *keyword*-analyse liet bovendien een aanvullend verschil zien: als Vlaamse recensenten de politiek-maatschappelijke eigenschappen van romans bespreken, spreken zij significant vaker over engagement, terwijl Nederlandse recensenten in dergelijke evaluaties eerder refereren aan de actualiteit. Naast de herkomst van de recensent bleek ook de herkomst van de auteur een factor van betekenis: zowel Nederlandse als Vlaamse critici hebben bij Vlaamse auteurs significant meer aandacht voor de ideologische eigenschappen dan bij Nederlandse auteurs – al is dat verschil bij Vlaamse critici aanzienlijk groter.

Hoewel er in Vlaamse recensies dus meer aandacht uitgaat naar de ideologische eigenschappen, zagen we, over het gehele corpus genomen, niet alle autonomistische criteria ook vaker voorkomen in Nederlandse recensies en alle meer heteronomistische aspecten en eigenschappen in Vlaamse recensies. Stijlevaluaties komen in Nederlandse recensies weliswaar vaker voor, en auteursevaluaties in Vlaamse recensies, maar het aandeel van het aspect *structuur*, bijvoorbeeld, bleek in Vlaamse recensies juist significant hoger dan in Nederlands recensies. Ook de 'structurele' eigenschappen *samenhang* en *continuïteit* komen vaker voor in Vlaamse recensies, terwijl evaluaties die betrekking hebben op *humor* en *emotionaliteit* – twee meer affectieve eigenschappen – vaker voorkomen in Nederlandse recensies.[235] Dat Vlaamse recensenten dus romans politieker en meer ideologisch beoordelen dan hun Nederlandse collega's, betekent dus niet noodzakelijk dat ze daardoor ook minder aandacht hebben voor andere aspecten en eigenschappen die bij uitstek met strikt-autonomistische literatuurkritiek worden geassocieerd.

Dat het aandeel van *emotionaliteit* en *humor* in Nederlandse recensies hoger is, contrasteert bovendien met het idee dat de Vlaamse literatuur meer emotioneel is, terwijl de Nederlandse literatuur afstandelijker zou zijn. Tegelijkertijd heb ik mij in deze scriptie alleen geconcentreerd op dagbladrecensies van fictioneel proza. Nader onderzoek zou moeten uitwijzen of we dezelfde resultaten vinden wanneer we ook naar bijvoorbeeld tijdschriftrecensies of poëzierecensies kijken.[236] Voor tijdschriftrecensies van fictioneel proza verwacht ik dat de in deze scriptie gepresenteerde methodiek relatief goed in staat zal zijn evaluatieve uitspraken te classificeren, en er hoogstens een uitbreiding van de handmatig gecodeerde trainingsset nodig is. Voor poëzierecensies vermoed ik daarentegen dat wezenlijk andere aspecten en eigenschappen van belang zijn, waardoor een soortgelijke methode voor poëzierecensies min of meer vanaf de grond opgebouwd zou moeten worden, en het categoriseringsmodel van Linders en Op de Beek eigenlijk alleen als een globaal sjabloon kan dienen. Een verdere validatie van de modellen op andere genres en publicatiemedia zou dat echter definitief moeten uitwijzen.

De verwachting dat buitenliteraire aspecten langer hun stempel blijven drukken op de oordelen van Vlaamse critici, en werkimmanente criteria pas later toenemen, werd deels bevestigd. In Nederlandse recensies neemt de

235 De term 'structurele eigenschappen' is afkomstig van Sven Vitse. Zie: Vitse, "Esther op de Beek, Een literair fenomeen van de eerste orde. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: Een kwantitatieve en kwalitatieve analyse.", 149.

236 Zeker ook omdat Brems zich in zijn onderscheid tussen de emotionele Vlaamse literatuur en de afstandelijke Nederlandse literatuur voornamelijk concentreert op poëzie. Zie: Brems, "De relatie tussen de Nederlandse en de Vlaamse literatuur. De teller en de noemer".

aandacht voor *thematiek* en de ideologische eigenschappen sinds 1946 gestaag af, terwijl het aandeel van *structuur* vanaf 1946 juist toeneemt. In Vlaamse recensies zien we van 1946 tot de jaren zestig eerst nog een toename van evaluaties die betrekking hebben *thematiek*, *didactische eigenschappen, morele eigenschappen, politiek-maatschappelijke eigenschappen* en *religieuze eigenschappen*. Pas later, midden jaren zestig, neemt het aandeel van de ideologische eigenschappen af, en in deze periode kunnen we ook de eerste duidelijke toename in het aandeel van structuurevaluaties zien.

Daarmee zien we inderdaad – op bepaalde aspecten en eigenschappen, althans – een latere afname van buitenliteraire factoren en een toename van werkimmanente criteria in Vlaamse recensies dan in Nederlandse recensies. Tegelijkertijd voltrekken de sterkste veranderingen zich eerder dan je op basis van conventionele periodiseringen van de ontzuiling in Vlaanderen en (tot op zekere hoogte) in Nederland zou verwachten. In Nederlandse recensies wordt de jaren zestig immers als het epicentrum van de ontzuiling aangewezen, terwijl de bovengenoemde ontwikkelingen zich dus al eerder aantekenen. Na de jaren zestig neemt het aandeel van *structuur* zelfs nog nauwelijks toe, en een *Merlyn*-effect kan in deze scriptie dan ook niet aannemelijk gemaakt worden. Alleen het aandeel van *stijl* neemt tijdens de jaren zestig relatief sterk toe, maar dat is ook het geval in Vlaamse recensies. In Vlaanderen wordt de ontzuiling vooral rond de jaren zeventig en tachtig gesitueerd, terwijl in de jaren zestig het aandeel van het aspect *structuur* (en dus ook *stijl*) al toeneemt, en het aandeel van *thematiek* en de ideologische eigenschappen afneemt. Wel neemt het aandeel van *structuur*, *stijl*, en *samenhang* in Vlaamse recensies in de jaren tachtig nog relatief sterk toe, en het aandeel van *auteur/intentie* relatief snel af, maar zeker voor *structuur* en *stijl* geldt dat deze toename minder sterk is dan in de jaren zestig. Bovendien verandert het aandeel van de ideologische eigenschappen in de jaren tachtig nauwelijks, terwijl 1985 volgens Hugo Brems wel een belangrijk kantelpunt zou markeren in de levensbeschouwelijke aard van het Vlaamse literaire veld.

De resultaten suggereren enerzijds dan ook dat tekenen van ontzuiling in het Vlaamse en Nederlandse literaire veld al eerder duidelijk zichtbaar zijn. Dat komt, zeker voor Nederland, ook niet geheel uit de lucht vallen: zoals ik in de inleiding al aangaf, hebben onder andere Anbeek en Brems beargumenteerd dat de ontzuiling van het Nederlandse literaire veld al ruim voor de jaren zestig was begonnen.[237] Sterker nog: volgens Anbeek kan 'het literaire leven' vlak na de oorlog 'met recht ontzuild worden genoemd'.[238] En ook in het geval van Vlaanderen erkent Brems dat hoewel 1985 een belangrijk kantelpunt markeert, 'de ideologische scheidslijnen in de literaire praktijk zelf' dan al 'verregaand vervaagd' zouden zijn, al biedt hij daarbij geen alternatieve periodisering.[239]

237 Anbeek, "Niet het moment voor experimenten. De twee gezichten van de jaren vijftig"; Brems, *Altijd weer vogels die nesten beginnen: geschiedenis van de Nederlandse literatuur, 1945-2005*, 36.

238 Anbeek, "Niet het moment voor experimenten. De twee gezichten van de jaren vijftig", 21.

239 Brems, *Altijd weer vogels die nesten beginnen: geschiedenis van de Nederlandse literatuur, 1945-2005*, 484.

Anderzijds heb ik laten zien dat de ontwikkeling van structuurevaluaties in Nederland ook per krant verschilt. Waar, over het hele corpus genomen, de sterkste toename van structuurevaluaties tussen 1945 en 1960 zichtbaar is, neemt de proportionele aandacht voor *structuur* in *de Volkskrant* wél het sterkst toe vanaf de tweede helft van jaren zestig. Voor *Algemeen/NRC Handelsblad* en *Trouw* zien we de sterkste toename zelfs tussen 1975 en 1985, terwijl de aandacht voor structuurevaluaties in *De Telegraaf* tussen 1965 en 1985 juist afneemt. Deze onderlinge verschillen wijzen erop dat de ontzuiling in literaire recensies mogelijk niet zozeer als één min of meer homogeen fenomeen geanalyseerd kan worden, maar dat we bij de verschillende zuilen in Nederland en Vlaanderen – die immers bij uitstek vertegenwoordigd worden door de verschillende dagbladen – op meerdere momenten in de tijd en op meerdere manieren de effecten van de ontzuiling kunnen waarnemen. Aanvullend onderzoek, waarin ook de ontwikkelingen en de historische posities van de verschillende zuilen in Nederland en Vlaanderen worden vergeleken, zou kunnen uitwijzen of er inderdaad sprake is van een blinde vlek in mijn analyse, om daarmee tot een secuurder begrip van het effect van de ontzuiling op literaire recensies te komen.

Ook de historische ontwikkeling van de media, en dan in het bijzonder de toenemende commercialisering van de dagbladkritiek, speelt hier mogelijk nog een rol van betekenis. Waar ik deze factoren in deze scriptie buiten beschouwing heb moeten laten, hebben Linders en Op de Beek literaire evaluaties in Nederlandse recensies wel vanuit een dergelijke invalshoek bestudeerd. Zij rapporteren daarbij 'onvoldoende' bewijs voor 'een patroon van commercialisering'.[240] Toch verdient dit mijns inziens nader onderzoek, onder andere omdat Linders en Op de Beek de krant waarvan tekenen van commercialisering wellicht nog het meest evident zichtbaar zouden kunnen zijn, *De Telegraaf*, volledig buiten beschouwing hebben gelaten. Zoals ik echter in deze scriptie heb laten zien, vertoont *De Telegraaf* wat betreft structuurevaluaties al markante tegendraadse trekken. Nader onderzoek zou kunnen uitwijzen of hier sprake is van een structureel patroon, hoe dat zich verhoudt tot de commercialisering van de dagbladkritiek, en of we in Vlaamse populaire kranten (zoals *Het Laatste Nieuws*) gelijkaardige patronen aantreffen.

Met de verschillen die we ook tussen kranten onderling vinden, raken we aan een andere kwestie waar ik in deze scriptie een subtiele balans in heb proberen te vinden: het interpreteren van de omvang van de gevonden verschillen. 'Nederlands' en 'Vlaams' bleken in deze scriptie geen irrelevante categorieën. Wanneer je deze variabelen willekeurig zou husselen, wanneer je zou pretenderen dat Jos Borré's recensie van *Het verdriet van België* geschreven had kunnen zijn door een Nederlander, en Willem Kuipers' recensie van *Red ons, Maria Montanelli* door een Vlaming, is dat de kans dat je gelijkaardige verschillen tussen Nederlandse en Vlaamse recensies aantreft zeer klein.[241] Dat betekent echter, zoals ik eerder heb benadrukt, niet noodzakelijk dat deze verschillen ook bijzonder betekenisvol zijn: met meer dan 20.000 observaties kunnen ook vrij kleine verschillen significant zijn. En inderdaad:

240 Linders, "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005", 250.

241 Husselmetafoor en voorbeelden geïnspireerd op: Jockers en Mimno, "Significant themes in 19th-century literature", 756.

op slechts drie van de achtentwintig aspect- en eigenschapscodes vonden we *geen* significante verschillen tussen Nederlandse en Vlaamse recensies.

Om de omvang van de verschillen tussen Nederlandse en Vlaamse recensies beter te kunnen duiden, rapporteerde ik naast de statistische significantie ook de effectgroottes van de verschillen. Bekeken vanuit de conventionele vuistregels van Jacob Cohen, waarbij een effectgrootte r vanaf 0.10 een klein effect, een r vanaf 0.30 middelgroot, en vanaf 0.50 een groot representeert, zouden we moeten concluderen dat we vooral kleine verschillen tussen Nederlandse en Vlaamse recensies aantreffen. De grootste effectgrootte die ik in deze scriptie heb gerapporteerd, was 0.22 – en daarmee niet hoog genoeg om in aanmerking te komen voor het label 'middelgroot'. Zoals ik echter in het zesde hoofdstuk van deze scriptie heb willen beargumenteren, geeft een dergelijke contextafhankelijke interpretatie van effectgroottes een vertekend beeld van de verhoudingen tussen Vlaamse en Nederlandse recensies. Dat we in deze scriptie geen effectgroottes groter dan 0.5 zouden aantreffen, lag immers ook in de lijn der verwachting – veel van de variantie is simpelweg te verklaren aan de hand van het besproken boek.

Eerder dan gebruik te maken van Cohen's vuistregels, koos ik in deze scriptie voor een contextueel-interpretatief kader om de effectgroottes te duiden. Concreet betekende dit dat ik de effectgroottes zo veel mogelijk probeerde te duiden in vergelijking met de reikwijdte van effectgroottes in gelijkaardige contexten, en dan in het bijzonder de effectgroottes die we bij paarsgewijze vergelijkingen tussen krantentitels aantreffen. Zo bleken we, op basis van 3.240 vergelijkingen, tussen krantentitels nauwelijks effectgroottes groter dan 0.20 aan te treffen. Vanuit dat perspectief zijn meerdere verschillen die ik in deze scriptie heb aangetroffen, en dan in het bijzonder die op de ideologische eigenschappen, wel degelijk als relatief groot te bestempelen.

Hoewel ik ervan overtuigd ben dat een dergelijke aanvliegroute een accurater beeld schetst van de omvang van de verschillen dan de contextonafhankelijke vuistregels van Cohen, is ze niet zonder problemen. De grens tussen een effectgrootte van 0.09 en 0.10, of een p-waarde van 0.051 en 0.05 mag dan tot op zekere hoogte arbitrair zijn, het biedt wel duidelijkheid. Dat is veel minder het geval wanneer je de effectgroottes eerder contextueel interpreteert. Wat betekent een effectgrootte van 0.06 wanneer de gemiddelde effectgrootte tussen kranten onderling 0.08 is? Hoe groot is een r van 0.10 in deze context? Waar effectgroottes die sterk afwijken van de reikwijdte nog relatief makkelijk zijn te duiden, zijn de waardes die daartussen liggen meer afhankelijk van de interpretatie van de onderzoeker, en daar schuilt – ook bij mij – het risico van *confirmation bias*.

Voor de particuliere problematiek van de vergelijking tussen Nederlandse en Vlaamse recensies zou verder onderzoek deels een uitkomst kunnen bieden. Waar ik mij deze scriptie beperkt heb tot de Nederlandse en Vlaamse literatuurkritiek, zouden we met een multinationale comparatieve analyse, waarin bijvoorbeeld ook de Duitse, Engelse en Franse literatuurkritiek vergeleken wordt, de omvang van de verschillen tussen Nederland en Vlaanderen mogelijk beter kunnen contextualiseren. Wanneer we Duitse met Nederlandse recensies vergelijken, of Vlaamse met Franse, zien we dan verschillen van dezelfde omvang? Een multinationaal comparatief perspectief

zou uiteraard niet alleen een verdere contextualisering van de omvang van de verschillen tussen Nederland en Vlaanderen opleveren, maar biedt ook de mogelijkheid om de verhoudingen tussen verschillende Europese literaturen en de invloed van nationaliteit op literaire oordelen beter te begrijpen. Recente computationeel-taalkundige doorbraken, zoals meertalige tekstclassificatie, waarin een model wordt getraind op één taal, maar ook accurate voorspellingen kan maken op teksten in andere talen, bieden voor een dergelijk onderzoek ook veelbelovende mogelijkheden.[242]

Tegelijkertijd is de problematiek rond de interpretatie van geesteswetenschappelijke kwantitatieve resultaten in het algemeen, en van effectgroottes in het bijzonder, niet op te lossen met individuele onderzoeken, maar is hier veel eerder sprake van een breder wetenschapsgemeenschappelijk vraagstuk. Er valt bijzonder veel te zeggen over p-waardes en bijbehorende significantieniveau's, maar het zijn wel standaarden die door een grote wetenschappelijk gemeenschap worden geaccepteerd. In de *Digital Humanities* zijn dergelijke standaarden en conventies aanzienlijk minder sterk ontwikkeld. Sterker nog: de toepassing van statistiek lijkt überhaupt niet zozeer de norm.[243] Daarmee wil ik allerminst, zoals Nan Z. Da lijkt te prefereren, pleiten voor het automatisch overnemen van 'harde' normen van de exacte of de sociale wetenschappen.[244] In tegendeel: het verschil tussen Nederlandse en Vlaamse recensies is iets wezenlijks anders dan de effectiviteit van een medicijn. Juist vanwege dat verschil zie ik een noodzaak voor breed gedragen richtlijnen voor de interpretatie van kwantitatieve resultaten in de *Digital Humanities*. Richtlijnen die uiteraard rekening houden met de context, met het type onderzoek, de vraagstelling, het meetinstrument.

In deze scriptie probeerde ik voor de statistische toetsing en het berekenen van gemiddeldes zo veel mogelijk de analysemethode van Linders en Op de Beek aan te houden – zij het met enkele uitbreidingen, zoals het rapporteren van effectgroottes. Daarvoor koos ik niet alleen, zoals ik eerder heb benadrukt, voor de vergelijkbaarheid, maar ook vanuit het idee dat een dergelijke analysemethode indicatief is voor de (minimale) statistische standaarden in ons vakgebied. Een nadeel van deze analysemethode is echter dat het niet structureel rekening houdt met verschillen die we ook tussen andere groepen aantreffen – met verschillen tussen kranten, auteurs, recensenten of genres. Een gelaagder statistisch model, waarin we ook dergelijke variabelen *structureel* in onze analyse verwerken, zou dan ook een preciezer begrip op kunnen leveren van de manier waarop nationaliteit – te midden van al deze andere verschillen – een rol speelt in recensies.

242 Zie: Guillaume Lample en Alexis Conneau, "Cross-lingual language model pretraining", *arXiv preprint arXiv:1901.07291*, 2019,

243 Om een voorbeeld te noemen uit een paper ik dat anderszins enorm waardeer: in een onderzoek naar de diversiteit van romanpersonages schrijven de onderzoekers: 'We will not substantiate our findings using statistical legitimations. (...) In our inclusive approach all results are "significant"'. Ik vermoed dat dit specifieke argument, waarbij het ontbreken van statistische validatie wordt gekoppeld aan het overwegend positief geconnoteerde inclusiviteit, in andere wetenschappelijke velden niet als valide geaccepteerd zou zijn. In: Deijl en andere, "Mapping the Demographic Landscape of Characters in Recent Dutch Prose: A Quantitative Approach to Literary Representation", 27.

244 Da, "The Computational Case against Computational Literary Studies".

Daarbij zou de onzekerheid van de classificatiemodellen evenzeer nog explicieter in de analyse verwerkt kunnen worden. Hoewel de in deze scriptie gebruikte modellen literaire evaluaties relatief accuraat konden classificeren, waren ze allerminst foutloos. Zeker het voorspellen van de eigenschapscodes bleek voor de computermodellen nog een behoorlijk uitdagende taak. En deze fouten hebben ongetwijfeld ook een effect op de uiteindelijke analyse. De mogelijkheid om de resultaten, voor Nederlandse recensies althans, met die van Linders en Op de Beek te vergeleken, bood hier nog enigszins een stok achter de deur. Daarbij kwamen mijn resultaten in grote lijnen overeen met die van Linders en Op de Beek: ook bij hen nam de aandacht voor bijvoorbeeld *auteur/intentie* en de ideologische eigenschappen vanaf 1955 gestaag af, de aandacht voor *plot* toe, en bleven labels als *geheel/deel*, *overig*, *setting* en *stijl* relatief stabiel. Radicaal tegenovergestelde trends, waarbij een label bij Linders en Op de Beek toeneemt en bij mij juist sterk afneemt, of andersom, ben ik niet tegengekomen. Maar voor de iets subtielere verschillen, daarentegen, zoals de eerdere toename van structuurevaluaties in Nederlandse recensies, blijft het moeilijk om definitief uit te sluiten of deze inderdaad het gevolg zijn van classificatiefouten, of dat andere factoren, zoals corpusselectie of corpusgrootte, hier een rol spelen.

Samenvattend kunnen we, desalniettemin, stellen dat wanneer een Nederlandse literatuurcriticus een roman beoordeelt, hij dat wel degelijk ook als Nederlander doet. Wanneer een Vlaamse literatuurcriticus een roman beoordeelt, doet hij dat ook als Vlaming. Niet door zachtjes De Vlaamse Leeuw te neuriën, maar door net iets meer aandacht te hebben voor de ideologische eigenschappen van deze roman, voor engagement, voor originaliteit en literariteit, voor de auteur. Maar hij doet dit niet alleen als Vlaming. Hij doet dit evenzeer als recensent voor *De Morgen*, als recensent levend in het jaar 1983. En misschien ook wel simpelweg – je zou bijna zeggen: gewoon – als een groot bewonderaar van Hugo Claus.

BIBLIOGRAFIE

Absillis, Kevin. "'From now on we speak civilized Dutch': the authors of Flanders, the language of the Netherlands, and the readers of A. Manteau". *Language and Literature* 18, nummer 3 (2009): 265–280.

Allington, Daniel, Sarah Brouillette en David Golumbia. "Neoliberal tools (and archives): A political history of digital humanities". *LA Review of Books*, 2016.

Anbeek, Ton. *Geschiedenis van de Nederlandse literatuur tussen 1885 en 1985*. De Arbeiderspers, 1990.

———. "Het Vlaamse verschil". *Dietsche Warande en Belfort* 2 (1996): 199–210.

———. "Niet het moment voor experimenten. De twee gezichten van de jaren vijftig". *Een stille revolutie? Cultuur en mentaliteit in de lange jaren vijftig*, 1997, 19–26.

Baelen, Carlo Van. *Begrensd literair grensverkeer. De zichtbaarheid van 150 literaire titels in Vlaanderen en Nederland.* Taalunie, 2014.

Bal, Mieke. "Narratology: Introduction to the Theory of Narrative", 2009.

Bekkering, Harry. "Veroverde traditie: de poëticale opvattingen van S. Vestdijk". Proefschrift, [Sl: sn], 1988.

Biber, Douglas, en Susan Conrad. *Register, Genre, and Style*. Cambridge University Press, 2009.

Boonstra, HT. "Van waardeoordeel tot literatuuropvatting". *De Gids* 142, nummer 4 (1979): 243–253.

Boot, Peter. "Distant reading Nederlandstalige online boekrespons". In *Achter de verhalen*. 2018.

Boswijk, JHW. "Er zal altijd wel iets over geschreven worden. De ontwikkelingen binnen de evaluaties in recensies aan de hand van het prozawerk van Remco Campert". Masterscriptie, Radboud Universiteit, 2017.

Brandellero, Amanda, en Olav Velthuis. "Reviewing art from the periphery. A comparative analysis of reviews of Brazilian art exhibitions in the press". *Poetics* 71 (2018): 55–70.

Brems, Hugo. "'"Rond dezelfde tijd in Vlaanderen..."'". In *Een of twee Nederlandse literaturen? Contacten tussen de Nederlandse en Vlaamse literatuur sinds 1830*, 25–34. Peeters, 2008.

Brems, Hugo. *Altijd weer vogels die nesten beginnen: geschiedenis van de Nederlandse literatuur, 1945-2005*. B. Bakker, 2006.

———. "De relatie tussen de Nederlandse en de Vlaamse literatuur. De teller en de noemer". *Tijdschrift over Nederlandse Letterkunde* 5, nummer 1 (1988): 34–38.

———. "Het centrum als periferie als centrum". *Dietsche Warande en Belfort* 141, nummer 2 (1996): 211–218.

———. "Het kritische discours over Vlaamse literatuur in Nederlandse kranten en weekbladen (1980-1994)". *Nederlandse Letterkunde* 4 (1999): 98–115.

———. "Nederlandse literatuur in Vlaanderen". *Literatuur* 93, nummer 1 (1993): 314–320.

Buelens, Geert. "Gotspe of godsgeschenk? Enkele notities over zin en onzin van een Belgische literatuurgeschiedenis". In *Literaire belgitude littéraire. Bruggen en beelden. Vues du Nord*, 67–80. Presses universitaires de Louvain, 2011.

———. "Naar een Vergelijkende Neerlandistiek". In *Colloquium Neerlandicum*. 2018. https://www.platformleest.org/lezing/naar-een-vergelijkende-neerlandistiek.

———. *Van Ostaijen tot heden: zijn invloed op de Vlaamse poëzie*. Vantilt, 2008.

Burgh, Benjamin van der, en Suzan Verberne. "The merits of Universal Language Model Fine-tuning for Small Datasets–a case with Dutch book reviews". *arXiv preprint arXiv:1910.00896*, 2019.

Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk en Yoshua Bengio. "Learning phrase representations using RNN encoder-decoder for statistical machine translation". *arXiv preprint arXiv:1406.1078*, 2014.

Chollet, François en andere. *Keras*. https://keras.io, 2015.

Cohen, J. *Statistical power analysis for the behavioral sciences. Abingdon*, 1988.

Da, Nan Z. "The Computational Case against Computational Literary Studies". *Critical Inquiry* 45, nummer 3 (2019): 601–639.

Dalen-Oskam, Karina van. "De stijl van R. (oratie)", 2013.

———. "Ballingschap der empirici". *Tijdschrift voor Nederlandse Taal-en Letterkunde* 125, nummer 2 (2011): 122–125.

De Sá, Cláudio Rebelo, Carlos Soares, Alipio Mário Jorge, Paulo Azevedo en Joaquim Costa. "Mining association rules for label ranking". In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 432–443. Springer, 2011.

Deijl, Lucas van der, Saskia Pieterse, Marion Prinse en Roel Smeets. "Mapping the Demographic Landscape of Characters in Recent Dutch Prose: A Quantitative Approach to Literary Representation". *Journal of Dutch Literature* 7, nummer 1 (2016).

Deneckere, Gita, Tom De Paepe en Bruno De Wever. *Een geschiedenis van België*, 2012.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee en Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding". *arXiv preprint arXiv:1810.04805*, 2018.

Dhingra, Bhuwan, Hanxiao Liu, Ruslan Salakhutdinov en William W Cohen. "A comparative study of word embeddings for reading comprehension". *arXiv preprint arXiv:1703.00993*, 2017.

Dierickx, Guido. "De verzuiling van belangengroepen: over de kunst van beschrijving en verklaring". In *Tussen bescherming en verovering: sociologen en historici over zuilvorming*, 41–65. 1988.

Dijk, N van, en S Janssen. "De reuzen voorbij. De metamorfose van de literaire kritiek in de pers sedert 1965". *J. Bardoel et al (red.), Journalistieke cultuur in Nederland, Amsterdam University Press, Amsterdam*, 2002, 209–235.

Dorleijn, Gillis J, en Cornelis Jacob van Rees. *De productie van literatuur. Het Nederlandse literaire veld 1800-2000*. Vantilt, 2006.

Durlak, Joseph A. "How to select, calculate, and interpret effect sizes". *Journal of pediatric psychology* 34, nummer 9 (2009): 917–928.

Eve, Martin Paul. *Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell's Cloud Atlas*. Stanford University Press, 2019.

Field, Andy, Jeremy Miles en Zoë Field. *Discovering statistics using R*. Sage publications, 2012.

Fleming, Paul. "Tragedy, for Example: Distant Reading and Exemplary Reading (Moretti)". *New Literary History* 48, nummer 3 (2017): 437–455.

Freund, Elizabeth. *Return Of Reader*. Routledge, 2013.

Gavin, Michael. "Is there a text in my data? (Part 1): On Counting Words". *Cultural Analytics*, 2019.

Gompel, Maarten van, en Martin Reynaert. "FoLiA: A practical XML format for linguistic annotation–a descriptive and comparative study". *Computational Linguistics in the Netherlands Journal* 3 (2013): 63–81.

Griswold, Wendy. "The fabrication of meaning: Literary interpretation in the United States, Great Britain, and the West Indies". *American Journal of Sociology* 92, nummer 5 (1987): 1077–1117.

Hayles, N Katherine. "How we read: Close, hyper, machine". *ADE bulletin* 150, nummer 18 (2010): 62–79.

Hellemans, Staf. "Verzuiling en ontzuiling van de katholieken in België en Nederland. Een historisch-sociologische vergelijking". *Sociologische Gids* 35, nummer 1 (1988): 43–56.

Humbeeck, Kris. "Een knuppel in het boekenhok: schrijf Vlaamse literatuur, geen Zuid-Nederlandse!" *De standaard*, 2004. https://www.standaard.be/cnt/dma21042005_021.

Huyse, Luc. "De ontzuiling is nog niet verteerd". *Samenleving en politiek* 20, nummer 2 (2013): 30–39.

Jannidis, Fotis. "On the perceived complexity of literature. A response to Nan Z. Da". *Cultural Analytics*, 2019.

Janssen, Susanne. *In het licht van de kritiek: variaties en patronen in de aandacht van de literatuurkritiek voor auteurs en hun werken*. Deel 14. Uitgeverij Verloren, 1994.

Janssens, Joris. "Bloeiende roos of aangestoken peer? Nut en onnut van cijfers inzake de Noord-Zuidkwestie". In *Een of twee Nederlandse literaturen? Contacten tussen de Nederlandse en Vlaamse literatuur sinds 1830*, 247–262. Peeters, 2008.

Jautze, Kim, Corina Koolen, Andreas van Cranenburgh en Hayco de Jong. "From high heels to weed attics: a syntactic investigation of chick lit and literature". In *Proceedings of the Workshop on Computational Linguistics for Literature*, 72–81. 2013.

Jockers, Matthew L. *Macroanalysis: Digital methods and literary history*. University of Illinois Press, 2013.

Jockers, Matthew L, en David Mimno. "Significant themes in 19th-century literature". *Poetics* 41, nummer 6 (2013): 750–769.

Joosten, Jos. "6 A Big Bang? The Paradox of the Reception of Günter Grass's Die Blechtrommel in The Netherlands". In *The Echo of Die Blechtrommel in Europe*, 120–135. BRILL, 2016.

———. "Opkomst en ondergang van *Merlyn* in de Nederlandse literatuurkritiek". In *Staande receptie: over literatuur, kritiek en literatuurwetenschap*, 59–68. Nijmegen: Vantilt, 2012.

———. *The Best Intentions: Literary Criticism in The Netherlands 1945-2005*. https://www.nwo.nl/onderzoek-en-resultaten/onderzoeksprojecten/i/10/3010.html. Geraadpleegd: 2019-05-08.

Jurafsky, Daniel, en James H. Martin. "Speech and Language Processing". Draftversie van 2 oktober 2019, beschikbaar via https://web.stanford.edu/~jurafsky/slp3/, 2019.

Karsdorp, Folgert. *Retelling Stories. A Computational-Evolutionary Perspective*. [Sl: sn], 2016.

Kennedy, James C. *Nieuw Babylon in aanbouw: Nederland in de jaren zestig*. Boom Koninklijke Uitgevers, 1995.

Kuzmenko, Audrey. *Gaussian Process (GP) smoothing*. https://docs.pymc.io/notebooks/GP-smoothing.html, 2018.

Lample, Guillaume, en Alexis Conneau. "Cross-lingual language model pretraining". *arXiv preprint arXiv:1901.07291*, 2019.

Leuker, Maria-Th. "Convergentie van Noord en Zuid na 1830? De briefwisseling tussen F.A. Snellaert en J.A. Alberdingk Thijm, 1843-1872". In *Een of twee Nederlandse literaturen? Contacten tussen de Nederlandse en Vlaamse literatuur sinds 1830*, 71–91. Peeters, 2008.

Linders, Yvette. "Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005". Proefschrift, Radboud Universiteit Nijmegen, 2014.

———. "TS Tools". *Tijdschrift voor Tijdschriftstudies*, nummer 32 (2012): 191–196.

Liu, Alan. "Where is cultural criticism in the digital humanities?", 2012, 490–509.

Liu, Liyuan, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao en Jiawei Han. "On the variance of the adaptive learning rate and beyond". *arXiv preprint arXiv:1908.03265*, 2019.

Long, Hoyt, en Richard Jean So. "Literary pattern recognition: Modernism between close reading and machine learning". *Critical Inquiry* 42, nummer 2 (2016): 235–267.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado en andere. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org, 2015. https://www.tensorflow.org/.

Media.(VRM), Vlaamse Regulator voor de. *Mediaconcentratie in Vlaanderen. Rapport 2019*, 2019.

Mikolov, Tomas, Kai Chen, Greg Corrado en Jeffrey Dean. "Efficient estimation of word representations in vector space". *arXiv preprint arXiv:1301.3781*, 2013.

Mooij, JJA. "De motivering van literaire waardeoordelen". *Tekst en lezer. Opstellen over algemene problemen van de literatuurstudie*, 1979, 253–278.

Moretti, Franco. "Conjectures on world literature". *New left review*, 2000, 54–68.

———. *Distant reading*. Verso Books, 2013.

———. *Graphs, maps, trees: abstract models for a literary history*. Verso, 2005.

Müller-Funk, Wolfgang. "Austrian Literature in a Trans-cultural Context". In *The architecture of modern culture: Towards a narrative cultural theory*, 3:248–263. Walter de Gruyter, 2012.

Musschoot, Anna. "Anbeeks keuze: enkele kanttekeningen". *Dietsche Warande en Belfort* 141, nummer 2 (1996): 233–242.

Nischik, Reingard M. "Margaret Atwood's Reception in Canada and the United States: A Comparative Analysis of North American Reviews of The Blind Assassin". *Contemporary Women's Writing* 11, nummer 3 (2017): 354–372.

Nischik, Reingard M. "Reviewing Atwood in Canada and the United States: From (Inter) Nationalism to Transnationalism". In *Comparative North American Studies*, 121–178. Springer, 2016.

Nooralahzadeh, Farhad, Lilja Øvrelid en Jan Tore Lønning. "Evaluation of Domain-specific Word Embeddings using Knowledge Resources". In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*. 2018.

Op de Beek, Esther. "'Een literair fenomeen van de eerste orde'. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: een kwantitatieve en kwalitatieve analyse". Proefschrift, Radboud Universiteit Nijmegen, 2014.

Pander Maat, Henk, Kay Raaijmakers, Dennis Vermeulen en Kees De Glopper. "Tekstkenmerken en tekstkwaliteit van leerlingteksten. Een annotatiestudie". *Tijdschrift voor Taalbeheersing* 41, nummer 2 (2019): 311–359.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel en andere. "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research* 12 (2011): 2825–2830.

Peng, Tony. "The Staggering Cost of Training SOTA AI Models". *Synced*, juni 2019. https://syncedreview.com/2019/06/27/the-staggering-cost-of-training-sota-ai-models/.

Peters, Matthew E, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee en Luke Zettlemoyer. "Deep contextualized word representations". *arXiv preprint arXiv:1802.05365*, 2018.

Pieterse, Saskia. "De buik van de lezer. Over spreken en schrijven in Multatuli's Ideeën", 2008.

Piper, Andrew. "Do we know what we are doing?" *Cultural Analytics*, 2019.

———. *Enumerations: Data and literary study*. University of Chicago Press, 2018.

Pires, Telmo, Eva Schlinger en Dan Garrette. "How multilingual is Multilingual BERT?" *arXiv preprint arXiv:1906.01502*, 2019.

Praamstra, Olf. "De analyse van kritieken". *Voortgang. Jaarboek voor de neerlandistiek* 5 (1984): 241–264.

Ramage, Daniel, David Hall, Ramesh Nallapati en Christopher D Manning. "Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora". In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, 248–256. Association for Computational Linguistics, 2009.

Ramsay, Stephen. *Reading Machines: Toward and Algorithmic Criticism*. University of Illinois Press, 2011.

Rees, CJ van. "De brugfunctie van Merlyn in de Nederlandse literatuurbeschouwing". *K. Fens & H. Verdaasdonk (red.), Op eigen gronden. Opstellen aangeboden aan Prof. Dr. JJ Oversteegen (...), Utrecht*, 1989, 43–64.

Rehurek, Radim, en Petr Sojka. "Software framework for topic modelling with large corpora". In *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Citeseer, 2010.

Righart, Hans. *De eindeloze jaren zestig. Geschiedenis van een generatieconflict*. De Arbeiderspers, 1995.

Risch, Julian, en Ralf Krestel. "Domain-specific word embeddings for patent classification". *Data Technologies and Applications* 53, nummer 1 (2019): 108–122.

Rosenthal, Jesse. "Introduction:"narrative against data"". *Genre* 50, nummer 1 (2017): 1–18.

Ruiter, Frans, en Wilbert Smulders. *Literatuur en moderniteit in Nederland: 1840-1990*. De Arbeiderspers, 1996.

Schäfer, Roland, en Felix Bildhauer. "Building large corpora from the web using a new efficient tool chain." In *LREC*, 486–493. 2012.

Schenkeveld-Van der Dussen, Maria A. *Nederlandse literatuur, een geschiedenis*. Uitgeverij Contact, 1998.

Schmuck, Lieselotte. "Literaturkritik und literarische Wertung. Aspekte einer inhaltsanalytischen Untersuchung deutschsprachiger Romankritik (1945-75)". *Zeitschrift fur Literaturwissenschaft und Linguistik, Beiheft* 12 (1981): 96–115.

Schöch, Christof. "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama." *DHQ: Digital Humanities Quarterly* 11, nummer 2 (2017).

Smeets, Roel, en Lucas van der Deijl. "Tussen close en distant. Personagehiërarchieën in Peter Buwalda's Bonita Avenue". *Tijdschrift voor Nederlandse Taal-en Letterkunde* 134 (2018): 123–145.

Smeets, Roel, Eric Sanders en Antal van den Bosch. "Character Centrality in Present-Day Dutch Literary Fiction". *DH Benelux journal*, 2019.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever en Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting". *The journal of machine learning research* 15, nummer 1 (2014): 1929–1958.

Stronks, Els. "De afstand tussen close en distant. Methoden en vraagstellingen in computationeel letterkundig onderzoek". *Tijdschrift voor Nederlandse Taal-en Letterkunde* 129, nummer 4 (2014): 205–214.

Ter Meulen, JM. "Een krant die nooit z'n ziel verkocht". Masterscriptie, Radboud Universiteit, 2018.

Tsoumakas, Grigorios, Ioannis Katakis en Ioannis Vlahavas. "Mining multilabel data". In *Data mining and knowledge discovery handbook*, 667–685. Springer, 2009.

Tulkens, Stéphan, Chris Emmery en Walter Daelemans. "Evaluating unsupervised Dutch word embeddings as a linguistic resource". *arXiv preprint arXiv:1607.00225*, 2016.

Twisk, JWR. "Inleiding in de toegepaste biostatistiek". *Houten, The Netherlands: Springer Media BV*, 2016.

Underwood, Ted. "A Genealogy of Distant Reading." *DHQ: Digital Humanities Quarterly* 11, nummer 2 (2017).

———. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019.

Van Croonenborch, Johan. "De prijs van het publieke debat: een bedrijfseconomische geschiedenis van de Vlaamse krantenpers, 1830-1980". Proefschrift, Ghent University, 2018.

Van den Akker, Wiljan. *Een dichter schreit niet: aspecten van M. Nijhoffs versexterne poetica*. Veen, 1985.

Van den Akker, Wiljan, en Gillis Dorleijn. "Consensus en conflict. Over de literatuurgeschiedschrijving van Nederland en Vlaanderen". *Nederlandse letterkunde* 4, nummer 2 (1999): 191–210.

Van Humbeeck, Ben. "'Het karakter van myten.'Literaire identiteitsvorming in de Vlaamse literaire kritiek (1980-1994)". *Neerlandistiek* 2007 (2007).

Van Humbeeck, Ben, en Floor Van Renssen. "Op voet van gelijkheid? De Noord-Zuidkwestie in de Nederlandse en de Vlaamse literaire kritiek, 1980-1995". In *Een of twee Nederlandse literaturen? Contacten tussen de Nederlandse en Vlaamse literatuur sinds 1830*, 263–290. Peeters, 2008.

Van Renssen, Floor. *'Lezer, er zijn ook Belgen!': interactie tussen de Nederlandse en Vlaamse literatuur via literaire kritiek en uitgeverij (1980-1995)*. Maklu, 2013.

Vanheste, Bert. "Van de eenheidsworst, verlos ons, Ton". *Dietsche Warande en Belfort* 141 (1996): 229–232.

VERBI Software. *MAXQDA 2018*. Berlijn, Duitsland: VERBI Software, 2017. https://www.maxqda.com.

Verboord, Marc, en Suzanne Janssen. "Informatieuitwisseling in het huidige Nederlandse en Vlaamse literaire veld. Mediagebruik en gelezen boeken door literaire lezers en bemiddelaars". In *Een of twee Nederlandse literaturen? Contacten tussen de Nederlandse en Vlaamse literatuur sinds 1830*, 307–321. Peeters, 2008.

Verhoeven, Gerard. "Bespreking van Linders, Y. (2014). Met waardering gelezen. Een nieuw analyse-instrument en een kwantitatieve analyse van evaluaties in Nederlandse literaire dagbladkritiek, 1955-2005. Proefschrift, Radboud Universiteit Nijmegen. 288 pp". *Tijdschrift voor taalbeheersing* 37 (2015): 102–108.

Vermeulen, Marieke. "De dagbladkritiek van vóór en ná Merlyn. Verschillen tussen de literatuurkritische praktijk in 1955 en 1975". *Tijdschrift voor Tijdschriftstudies*, nummer 29 (2011): 18–35.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski en andere. "SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python". *arXiv e-prints*, arXiv:1907.10121, juli 2019, arXiv:1907.10121. arXiv: 1907.10121 [cs.MS].

Vitse, Sven. *Een woestijn die de stad verpulvert. Marxisme, kritiek en representatie in poststructuralisme, Ander Proza en postmodern proza*. Vrije Universiteit Brussel, 2008.

———. "Esther op de Beek, Een literair fenomeen van de eerste orde. Evaluaties in de Nederlandse literaire dagbladkritiek, 1955-2005: Een kwantitatieve en kwalitatieve analyse." *Tijdschrift voor tijdschriftstudies* 36 (2014): 147–152.

Vries, Wietse de, Andreas van Cranenburgh, Arianna Bisazza, Tommaso Caselli, Gertjan van Noord en Malvina Nissim. *BERTje: A Dutch BERT Model*, 2019. arXiv: 1912.09582 [cs.CL].

Witte, Els de, Dirk Luyten en Alain Meynen. *Politieke geschiedenis van België: van 1830 tot heden*. E-bookeditie. Manteau, 2016.

Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac en andere. "HuggingFace's Transformers: State-of-the-art Natural Language Processing". *ArXiv* abs/1910.03771 (2019).

Xiao, Han. *bert-as-service*. https://github.com/hanxiao/bert-as-service, 2018.

Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov en Quoc V Le. "XLNet: Generalized Autoregressive Pretraining for Language Understanding". *arXiv preprint arXiv:1906.08237*, 2019.

Hoofdstuk A

Bijlage

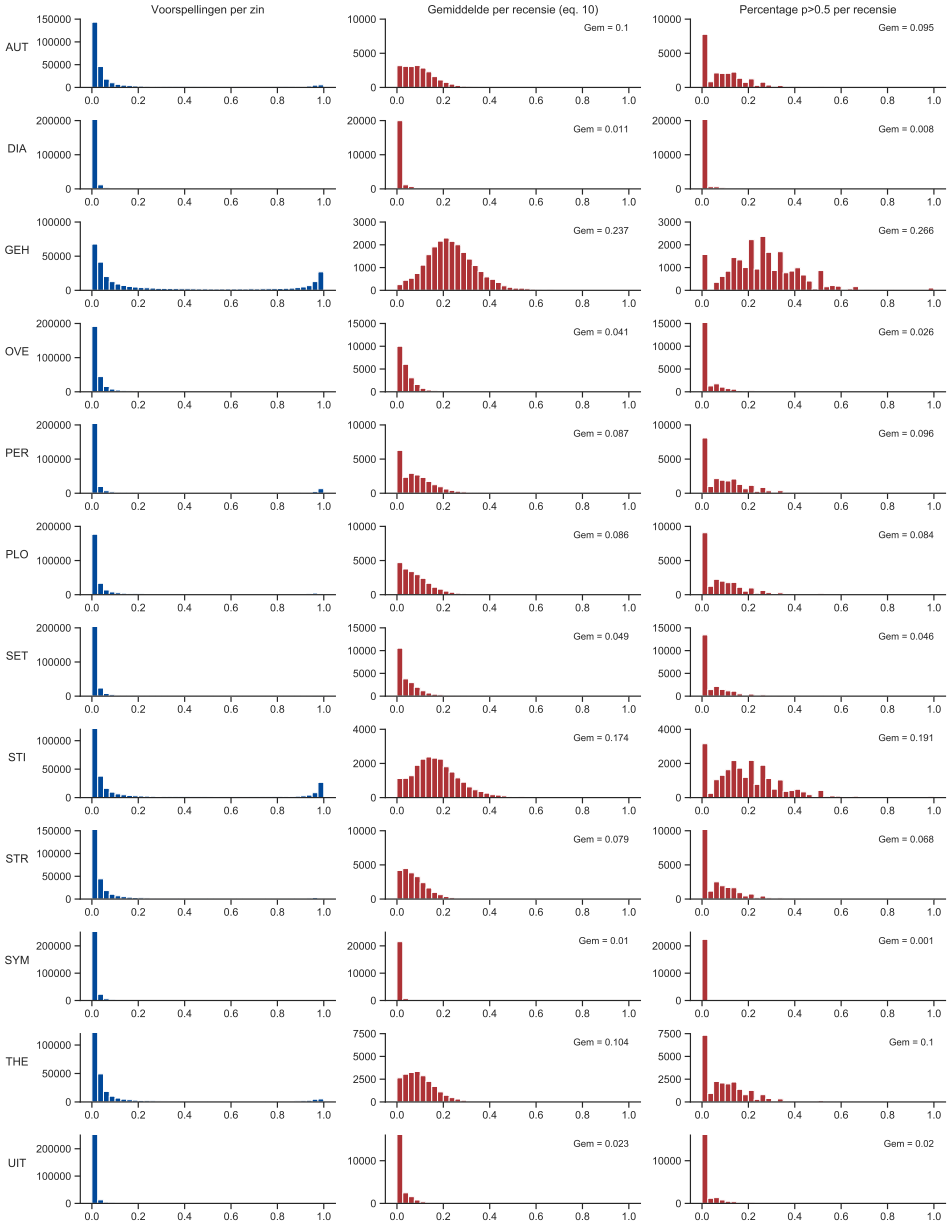
A.1 Eigenschapscodes per krant

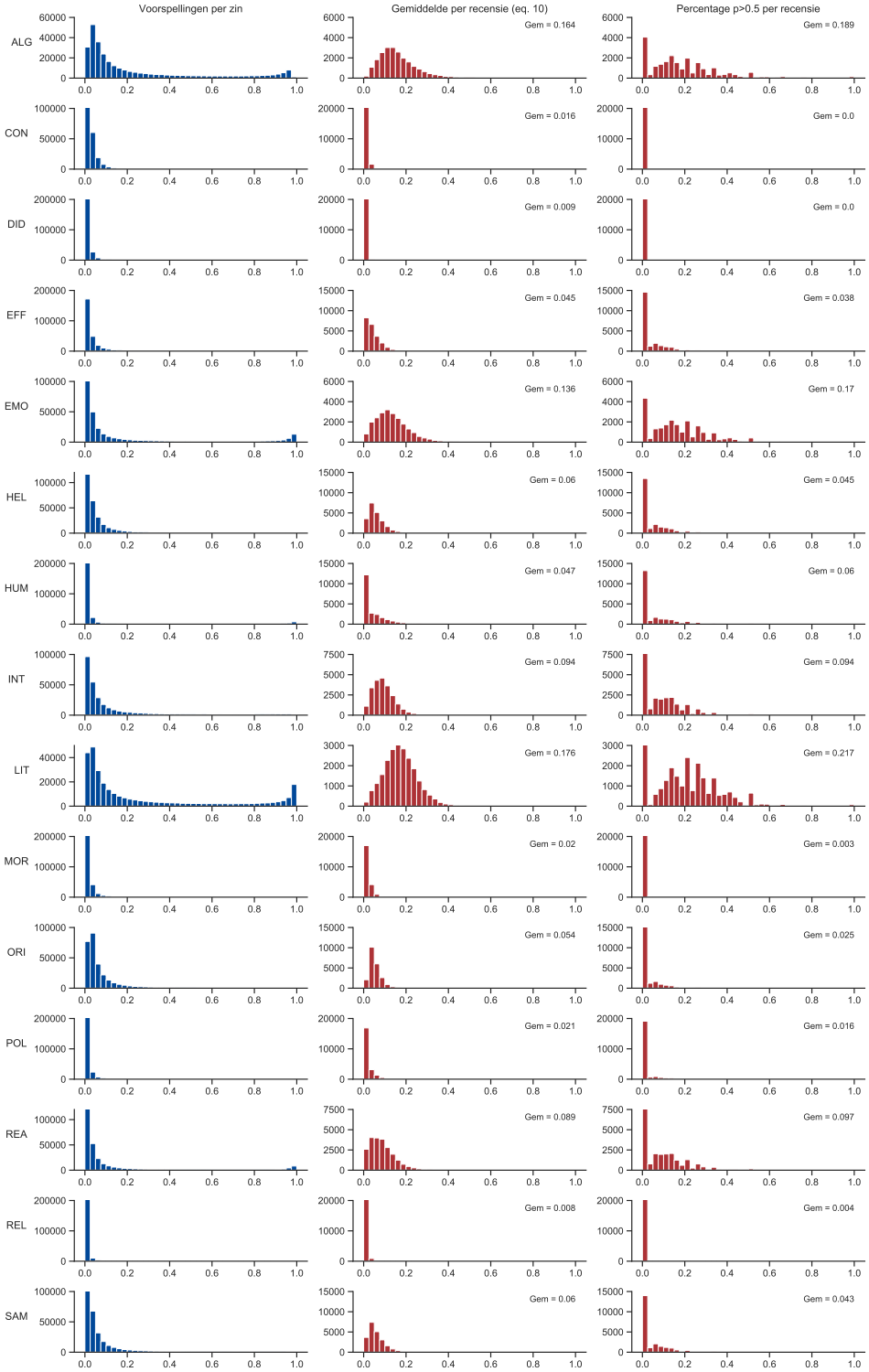
	Algemeen/overig	Continuiteit	Didactische eigenschappen	Efficiëntie	Emotionaliteit	Helderheid	Humor	Intensiteit/levendigheid	Literaire/intellectuele eigenschappen	Morele eigenschappen	Originaliteit	Politiek-maatschappelijke eigenschappen	Realiteitsgehalte/geloofwaardigheid	Religieuze eigenschappen	Samenhang
Algemeen Dagblad	20	1.6	0.9	5.1	14	5.6	5.7	8.5	16	1.8	5	1.9	8.3	0.76	5.2
Algemeen/NRC Handelsblad	16	1.6	0.98	4.8	14	6.3	5.1	9.7	17	2.1	5.4	2.2	8.9	0.79	6.2
De Gazet	10	1.1	1	3	18	2.5	14	9.8	19	2.2	4.7	1.2	8.7	1.1	2.7
De Morgen	15	1.6	0.91	4.6	14	5.6	4.8	10	18	2.1	5.7	2.2	8.3	0.72	5.7
De Nieuwe Gazet	16	1.7	1.1	4.1	11	5.8	3.8	9	21	2.1	5.4	2.5	9.4	0.76	6
De Nieuwe Gids	12	1.7	1.1	4.3	13	5.7	2.9	11	18	3.1	4.6	3.2	12	0.93	6.5
De Pers	16	1.2	0.87	2.8	15	3.8	7.9	8.9	20	2.3	6	2.5	9	0.56	3.8
De Standaard	15	1.7	0.95	4.6	13	5.8	4.3	9.4	19	2.1	5.8	2.3	8.8	0.74	6.3
De Telegraaf	17	1.5	0.9	4.2	15	5.2	5.3	9.6	17	1.9	5.1	2	8.7	0.68	5.4
De Tijd	17	1.6	0.87	4.9	14	5.1	4.8	9	20	1.9	5.4	2	8.1	0.65	5.3
De Tijd (NL)	16	1.8	1	4.6	12	6.4	4.8	8.7	18	2	5.4	2.1	8.9	0.83	7.4
De Volksgazet	15	1.8	1.2	3.1	15	5.5	3.6	9.3	21	2.7	4.9	3.1	6.9	1.1	6.5
De Waarheid	16	1.8	1.1	4.5	11	7	3.5	7.9	19	2.2	5.9	3.3	9.2	0.86	7.2
Gazet van Antwerpen	15	1.8	1.1	4.5	12	6	2.3	9.4	22	2.2	5.5	2.2	8.8	0.87	6.1
Het Belang van Limburg	19	1.9	0.96	5.3	11	5.4	2.6	9.1	19	1.9	6.3	2	8.2	0.73	6.7
Het Financieele Dagblad	15	1.5	0.93	4.1	17	5.3	5.5	10	16	2.2	5.1	2.3	8.7	0.72	5.6
Het Handelsblad	25	1.8	1	0.93	11	4.7	0.58	5.7	23	1.9	6.9	1.6	8.3	2.1	5.7
Het Laatste Nieuws	12	1.6	1.3	3.7	13	6	5.2	8.6	19	3.3	4.4	4.5	10	1.2	6
Het Nieuws van den Dag	17	2	1.1	7.2	13	6.1	1.8	12	19	2.8	4.2	1.6	5.6	0.92	6.4
Het Parool	19	1.6	0.89	4.6	13	5.9	5.1	8.9	17	1.8	5.4	1.7	8.7	0.64	5.6
Het Volk (VL)	17	1.6	0.93	3.7	11	5.8	3.4	12	23	1.9	5.9	1.8	7.2	0.76	4
Het Vrije Volk	19	1.5	0.96	4.4	13	5.6	4.6	8.7	19	2.1	5.2	1.8	9	0.74	4.8
Metro	25	1.6	0.81	5.1	13	6	2.3	7	16	1.7	6.2	1.4	7.3	0.42	5.5
Metro (NL)	47	1.8	0.55	0.77	6.8	3	0.61	1.6	5.6	1.7	8.3	2.6	2.3	0.49	17
NRC Next	16	1.4	0.97	4	16	5.6	5.5	9.2	18	2.1	5.4	2.4	8.6	0.76	4.6
Nederlands Dagblad	17	1.5	0.97	4.6	15	6.3	3.7	8.9	16	2.1	4.8	2.1	9.6	1.3	5.8
Reformatorisch Dagblad	18	1.6	1.2	4	16	6.3	2.7	7.6	16	2.7	4.6	2.6	9.2	2.3	5.9
Spits	26	1.2	0.82	3	11	5.8	6.5	6.6	12	1.9	6.5	1.9	13	0.6	4.1
Trouw	16	1.6	0.96	4.2	14	6.4	4.7	9.6	17	2	5.5	2.2	9.2	0.85	6.2
Volkskrant	16	1.6	0.9	4.6	13	6	4.8	9.8	17	1.9	5.4	2	9	0.73	6.3
Vooruit	15	1.9	1	3.5	13	5.1	5.6	8.1	20	2.1	6.7	1.6	11	0.8	6

A.2 TOP VEERTIG MEEST VOORKOMENDE WOORDEN IN POLITIEKE EVALUATIES

	Vlaamse politieke evaluaties			Nederlandse politieke evaluaties		
	Woord	Norm. freq.	χ^2	Woord	Norm. freq.	χ^2
1	politiek	128.09	-0.86	politiek	137.92	0.86
2	boek	99.24	1.23	roman	98.74	1.38
3	roman	88.28	-1.38	boek	89.53	-1.23
4	maatschappelijk	73.86	-0.44	maatschappelijk	79.34	0.44
5	verhaal	58.85	0.01	verhaal	57.99	-0.01
6	zullen	56.55	1.47	wel	51.92	6.75
7	schrijven	44.43	0.04	zullen	48.59	-1.47
8	engagement	42.12	12.60	maken	45.84	2.81
9	gaan	41.54	-0.00	moreel	43.69	0.66
10	tussen	40.39	3.66	schrijven	42.91	-0.04
11	sociaal	39.81	4.64	gaan	42.32	0.00
12	mens	39.81	0.11	actueel	40.95	8.08
13	moreel	38.66	-0.66	goed	40.55	3.54
14	waarin	36.93	-0.27	waarin	40.16	0.27
15	maken	35.77	-2.81	laten	39.18	1.21
16	wel	35.77	-6.75	schrijver	38.99	3.39
17	willen	34.04	-0.04	komen	38.59	7.90
18	laten	32.89	-1.21	mens	37.62	-0.11
19	leven	31.74	-0.88	geven	37.42	9.58
20	samenleving	31.16	1.25	leven	37.03	0.88
21	ander	30.58	0.01	moeten	35.85	1.13
22	zoals	30.00	-0.10	willen	35.46	0.04
23	goed	30.00	-3.54	zien	32.52	1.92
24	moeten	30.00	-1.13	zoals	31.93	0.10
25	auteur	29.43	11.65	ander	30.56	-0.01
26	schrijver	28.85	-3.39	tussen	30.37	-3.66
27	groot	28.85	-0.00	groot	29.19	0.00
28	eigen	27.12	0.97	sociaal	28.80	-4.64
29	heel	26.54	0.27	jaar	28.02	5.46
30	personage	26.54	4.25	wereld	28.02	1.59
31	lezer	25.97	0.00	samenleving	25.66	-1.25
32	actueel	25.39	-8.08	actualiteit	25.47	8.95
33	zien	25.39	-1.92	lezer	25.27	-0.00
34	komen	23.66	-7.90	engagement	24.88	-12.60
35	waar	23.08	-0.01	krijgen	24.49	7.85
36	tijd	22.50	-0.01	nieuw	24.10	0.48
37	maatschappij	22.50	2.49	waar	23.90	0.01
38	vinden	21.93	2.39	heel	23.90	-0.27
39	wereld	21.93	-1.59	vooral	22.92	1.29
40	staan	21.35	-0.00	manier	22.73	0.06

A.3 DISTRIBUTIES VAN WAARSCHIJNLIJKHEDEN




Voorspellingen per zin
Gemiddelde per recensie (eq. 10)
Percentage p>0.5 per recensie
ALG
CON
DID
EFF
EMO
HEL
HUM
INT
LIT
MOR
ORI
POL
REA
REL
SAM
Gem = 0.164
Gem = 0.189
Gem = 0.016
Gem = 0.0
Gem = 0.009
Gem = 0.0
Gem = 0.045
Gem = 0.038
Gem = 0.136
Gem = 0.17
Gem = 0.06
Gem = 0.045
Gem = 0.047
Gem = 0.06
Gem = 0.094
Gem = 0.094
Gem = 0.176
Gem = 0.217
Gem = 0.02
Gem = 0.003
Gem = 0.054
Gem = 0.025
Gem = 0.021
Gem = 0.016
Gem = 0.089
Gem = 0.097
Gem = 0.008
Gem = 0.004
Gem = 0.06
Gem = 0.043