

Publisher: Igitur, Utrecht Publishing & Archiving Services. Website: www.tijdschriftstudies.nl

Content is licensed under a Creative Commons Attribution 3.0 License

URN:NBN:NL:UI:10-1-113684. TS > # 32, december 2012, p. 191-196.

TS Tools

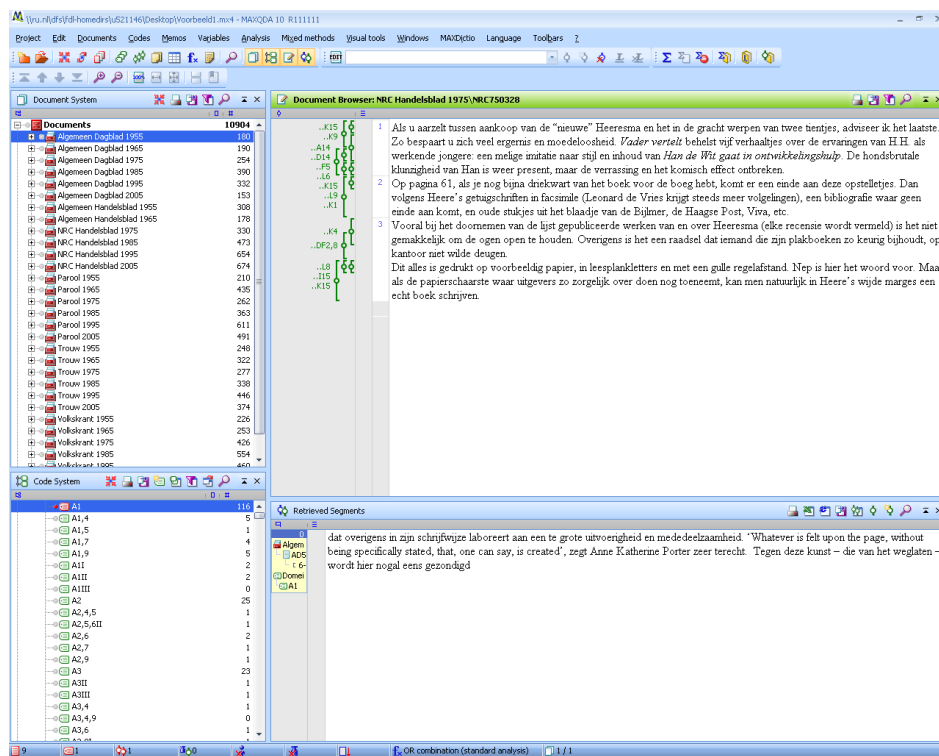
In deze aflevering van TS Tools laat promovenda Yvette Linders (Radboud Universiteit Nijmegen) zien hoe software voor kwalitatieve data-analyse kan worden toegepast in het onderzoek naar literatuurkritiek.

THE BEST INTENTIONS

Binnen het NWO-programma *The Best Intentions: Literary Criticism in the Netherlands 1945-2005* doe ik samen met Esther Op de Beek onderzoek naar de naoorlogse literatuurkritiek in Nederlandse dagbladen. We willen te weten komen hoe prozarecensies eruitzagen in de genoemde periode, of er verschillen tussen kranten zijn en of er sprake is van ontwikkelingen in de loop van de tijd. We zijn daarbij vooral geïnteresseerd in de evaluatieve passages en de aspecten en eigenschappen die daarin worden beoordeeld. Daarbij gaat het om aspecten als de stijl van het boek, de personages of de plot. Eigenschappen zijn bijvoorbeeld helderheid, levendigheid of emotionaliteit. Critici kunnen verschillende eigenschappen toekennen aan verschillende aspecten van een boek. Zo kan een criticus de stijl van een boek helder noemen, maar hij kan ook de personages helder vinden.

In ons onderzoek hebben we 734 recensies geanalyseerd om te achterhalen welke aspecten en eigenschappen veel voorkomen en welke minder van belang lijken te zijn. Het is onmogelijk om te werken met een corpus van deze omvang zonder software te gebruiken. Daarom zijn we op zoek gegaan naar een programma waarin we niet alleen alle recensies doorzoekbaar konden maken, maar ook op een overzichtelijke manier bij elke evaluatieve uitspraak konden aangeven welk aspect en welke eigenschap aan bod kwamen. Uiteindelijk hebben we gekozen voor MAXqda, een programma voor kwalitatieve data-analyse. Doorslaggevend in onze keuze waren de compatibiliteit met andere programma's, de zoekmogelijkheden en de stabiliteit van het programma. In MAXqda kunnen alle recensies in één bestand worden ondergebracht en vervolgens kunnen die recensies worden gecodeerd. De recensies zijn (als .doc(x)-, .rtf-, of .pdf-bestand) eenvoudig te laden in het programma.

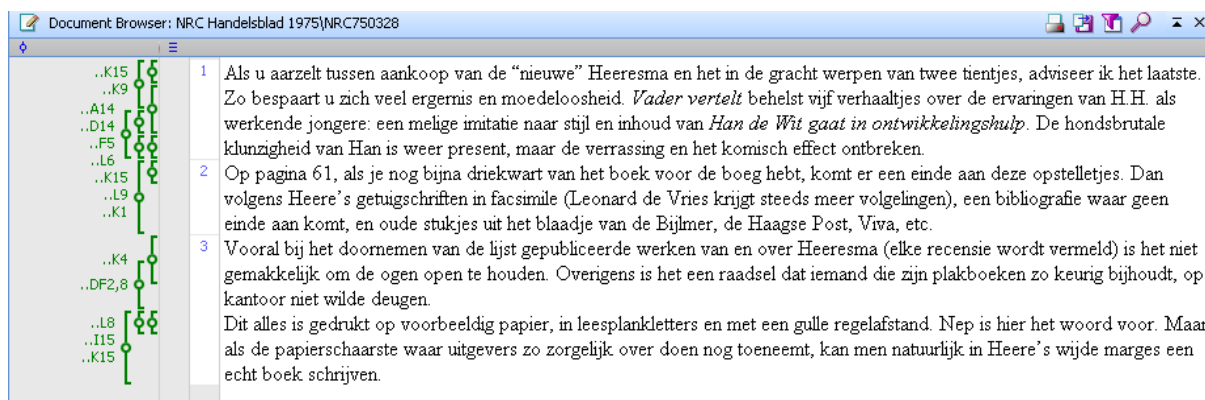
In de 734 recensies hebben we eerst de evaluatieve uitspraken gemarkeerd. Vervolgens hebben we elk van die uitspraken gecodeerd. Dat betekent dat voor elke evaluatieve uitspraak in de recensies is aangegeven welke combinatie van aspecten en eigenschappen werd besproken in die passage. In MAXqda komt dat er als volgt uit te zien:



Afb. 1: vensterstructuur MAXqda

VENSTERSTRUCTUUR

MAXqda deelt het scherm in vier verschillende vensters in. Linksboven is een overzicht van het volledige corpus te zien in het 'Document system'. Alle bestanden die zijn geïmporteerd zijn hier terug te vinden. Door op een document te klikken, wordt dat bestand opgeroepen in het veld rechtsboven, de 'Document browser'. Hierin is een tekst nog te bewerken. Bovendien is zichtbaar welke codes aan de tekst zijn toegekend.



Afb. 2: gecodeerde recensie in de 'Document Browser'

Onder het 'Document system', linksonder, staat het 'Code system', een overzicht van alle codes of categorieën. Het prettige hiervan is dat het te exporteren is, zodat er met een ander corpus opnieuw mee gewerkt kan worden. Wanneer een vergelijkbaar onderzoek gedaan zou worden naar bijvoorbeeld recensies in tijdschriften, kan een dergelijk

corpus worden aangelegd en met deze zelfde codes in MAXqda geanalyseerd worden.

In ons geval staan in het ‘Code system’ alle mogelijke combinaties van aspecten en eigenschappen die in het oordeel van de criticus aan bod kunnen komen. Het vierde veld in dit scherm, rechtsonder, bevat ‘Retrieved segments’. Resultaten van zoekopdrachten zullen hier verschijnen. Wanneer in het ‘Document system’ een groep recensies wordt geactiveerd en in het ‘Code system’ een code (dus in dit geval een combinatie van een aspect en een eigenschap), dan is in het veld ‘Retrieved segments’ te zien waar die code in de gekozen teksten voorkomt. Op die manier is bijvoorbeeld een overzicht op te roepen van alle passages van recensies uit het *Algemeen Dagblad* van 1955 waarin de efficiëntie van de stijl besproken wordt (code A1 in ons codesysteem).

Dankzij deze overzichtelijke indeling en de andere opties van MAXqda, konden we de vragen over ons corpus op een gestructureerde manier beantwoorden. We wilden zowel op kwantitatief als kwalitatief gebied meer te weten komen over de recensies.

KWANTITATIEVE ANALYSE

Hoewel MAXqda een programma is voor kwalitatieve data-analyse, is het zeker ook bruikbaar in onderzoek met een kwantitatieve component. In ons onderzoek is eerst een kwantitatieve analyse uitgevoerd. We waren zowel geïnteresseerd in het totaal aantal evaluaties in recensies in verschillende jaren en kranten als in de mate waarin verschillende aspecten en eigenschappen in (delen van) het corpus voorkwamen.

Om gegevens van verschillende kranten en jaren goed te kunnen scheiden, moeten de recensies in het ‘Document system’ gegroepeerd worden in mappen en sets. We hebben mappen gemaakt met daarin per map de recensies van een jaargang van een bepaalde krant (dus bijvoorbeeld alle recensies in het corpus uit het *Algemeen Dagblad* van 1955). Van de verschillende mappen kunnen vervolgens ‘sets’ gemaakt worden, bijvoorbeeld van alle recensies uit 1955 of alle recensies uit *De Volkskrant*. Daardoor is het niet alleen mogelijk om een specifieke jaargang van een krant nader te bestuderen, maar ook om een jaargang voor alle kranten samen te bekijken of een krant door de jaren heen te volgen.

Per document, per map, per set en in totaal wordt in het ‘Document system’ aangegeven hoeveel codes er zijn toegekend. Deze gegevens zijn later bruikbaar voor een uitgebreidere statistische analyse.

Hiernaast is te zien dat aan alle recensies samen in totaal 10904 codes zijn toegekend. Dit betekent dat er in het hele corpus 10904 evaluaties zijn gevonden. Daarvan zijn er 180 te vinden in recensies

Document	Count
Algemeen Dagblad 1955	180
AD550115	11
AD550305	38
AD550325	15
AD550924	20
AD551008	10
AD551029	15
AD551111	14
AD551126	12
AD551203	45
Algemeen Dagblad 1965	190
AD650213.1	11
AD650213.2	15
AD650406	16
AD650504	27
AD650529.1	8
AD650529.2	12
AD650604	15
AD650618	13
AD650630	12
-----	-
Totaal	10904

Afb. 3: groepering van recensies in het ‘Document System’

uit het *Algemeen Dagblad* in 1955. Op deze manier ontstaat al een eerste beeld van het aantal evaluatieve passages in de recensies. Zo is duidelijk te zien dat de aantallen evaluaties in recensies behoorlijk uiteenlopen: in de recensies die op het overzicht in de afbeelding te zien zijn, komen minstens 8 evaluaties voor; de meeste recensies bevatten tussen 10 en 20 evaluaties, maar er zijn in het *Algemeen Dagblad* in 1955 twee uitschieters met 38 en 45 evaluaties. Dit verschil in aantal evaluaties kan te maken hebben met verschillen in evaluatielengte. MAXqda kan, via een kleine omweg, het totaal aantal woorden van een recensie weergeven, maar andere programma's geven de resultaten overzichtelijker weer en zijn vaak duidelijker over waar woordgrenzen liggen. Wij hebben er daarom voor gekozen om de informatie over recensielengte, die in de kwantitatieve analyse niet mag ontbreken, niet uit MAXqda maar uit Word te halen.

Naast deze gegevens over het totaal aantal evaluaties is het ook mogelijk om kwantitatieve gegevens over bepaalde aspecten en eigenschappen op te roepen. Zoals gezegd bieden programma's voor kwalitatieve data-analyse de mogelijkheid om alle delen van documenten te verzamelen waaraan een bepaalde code is toegekend. Dat betekent dus dat het, wanneer alle documenten eenmaal gecodeerd zijn, mogelijk is om een overzicht te krijgen van alle passages waarin geoordeeld wordt over een bepaald aspect. MAXqda geeft na het activeren van alle codes met betrekking tot stijl een lijst van alle evaluatieve opmerkingen die betrekking hebben op dit aspect in het gehele corpus van 734 teksten. Op die manier wordt bijvoorbeeld al snel duidelijk dat van de bijna 11000 evaluaties in het corpus er ruim 2400 betrekking hebben op stijl. Ook specifiekere zoekopdrachten, bijvoorbeeld naar het aantal evaluaties met betrekking tot de helderheid van de stijl, kunnen worden uitgevoerd. Door een ander deel van de codes te activeren wordt de zoekopdracht verfijnd. Bovendien zijn ingewikkeldere zoekopdrachten mogelijk waarbij een overzicht wordt gevraagd van alle gevallen waarin een opmerking over dialogen voorkomt in de buurt van een opmerking over personages.

Deze gegevens kunnen niet alleen voor het gehele corpus verzameld worden, maar ook voor een bepaalde krant, voor een jaartal of voor combinaties van krant en jaartal. De kwantitatieve gegevens die MAXqda op deze manier levert (bijvoorbeeld over het aantal evaluaties over de helderheid van de stijl in verschillende jaren) zijn eenvoudig te exporteren naar SPSS of Excel voor verdere kwantitatieve analyse.

Op deze manier leverde de MAXqda-analyse, na een aantal eenvoudige bewerkingen in SPSS, gegevens op over welke aspecten en eigenschappen relatief gezien het meest en het minst besproken worden in recensies. Daarnaast kon statistisch getoetst worden tussen welke kranten en welke jaren er verschillen te zien waren in het bespreken van bepaalde aspecten en eigenschappen.

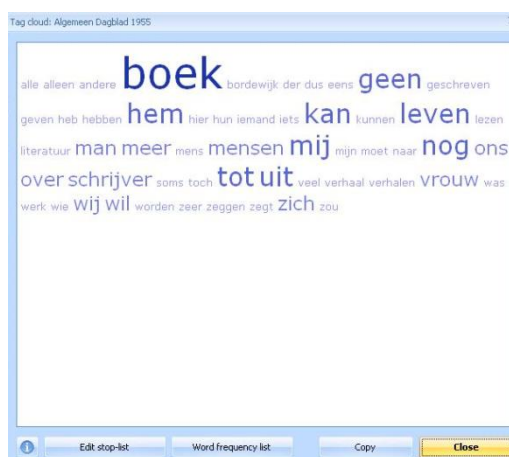
KWALITATIEVE ANALYSE

Wanneer we alleen naar de kwantitatieve analyse zouden kijken, zouden we de zaak simpeler voorstellen dan ze in feite is. Soms lijkt er op basis van een kwantitatieve analyse geen sprake van verschillen tussen kranten of ontwikkelingen in de tijd als het gaat om het beoordelen van een bepaald aspect of een bepaalde eigenschap, terwijl er inhoudelijk

nogal wat veranderd is. Zo wordt het aspect ‘stijl’ in alle jaren en in alle kranten relatief veel besproken, maar dat wil niet zeggen dat evaluaties van de stijl van een boek niet veranderd kunnen zijn in de loop van de tijd. Om daar meer over te weten te komen, is een uitgebreidere kwalitatieve analyse nodig. De lijsten die MAXqda maakt van bijvoorbeeld alle evaluaties uit (een deel van) het corpus die betrekking hebben op stijl, kunnen ook voor deze kwalitatieve analyse gebruikt worden. Door deze opmerkingen te bestuderen kan er een compleet beeld worden gevormd van wat er, in een bepaalde periode of een bepaalde krant, over stijl werd gezegd. Op deze manier konden we achterhalen of in elke periode hetzelfde werd verlangd van de stijl van een boek. Door de uitspraken uitgebreider te bekijken, wordt duidelijk of bijvoorbeeld een heldere stijl in alle jaren en alle kranten als iets positiefs wordt gezien, of dat helderheid van stijl soms ook wordt afgekeurd.

We waren niet alleen geïnteresseerd in de aard van het oordeel, maar ook in het gebruik van bepaalde begrippen. Het is mogelijk om de geselecteerde passages te doorzoeken op het voorkomen van bepaalde begrippen. Zo kan gezocht worden naar het woord ‘schrijven’. Het is ook mogelijk om een aantal woorden tegelijk te zoeken, of te werken met zogenaamde wildcards: zoeken op ‘schrijv* OR schrijf*’, zal als resultaat een overzicht geven van alle woorden als ‘schrijver’, ‘schrijven’, ‘schrijft’ en ‘schrijfster’. Daarbij kan binnen de evaluaties gezocht worden, maar ook het doorzoeken van het gehele corpus is mogelijk. Om vervolgens meer te weten te komen over de context waarin het woord in een bepaald document gebruikt is, is het voldoende om in de zoekresultaten op het woord te klikken. Het juiste document wordt dan geopend en het gezochte woord wordt gemarkeerd.

Door te zoeken naar bepaalde begrippen kan het verdwijnen of opkomen ervan worden aangetoond. Zo is bijvoorbeeld te zien dat het woord ‘kostelijk’ in de peiljaren 1955 en 1965 een aantal keer voorkomt in oordelen in recensies (in de betekenis ‘waardevol’ of ‘profijtelijk’), terwijl dit woord in latere recensies niet meer voorkomt. Naast specifieke zoektochten naar bepaalde begrippen kunnen ook de woordfrequentielijsten worden gebruikt die MAXqda maakt van het corpus, een deel van het corpus of een document. Op basis van deze woordfrequentielijsten maakt MAXqda ook woordwolken, ‘tag clouds’, waarin de begrippen die het meest voorkomen groot zijn weergegeven en de minder frequente woorden wat kleiner. Woorden die niet van belang zijn voor de analyse, zoals lidwoorden, kunnen uit deze ‘tag clouds’ verwijderd worden, om de woordwolk zo interessant en duidelijk mogelijk te maken. Zo is in onderstaand voorbeeld te zien dat naast ‘boek’ en ‘schrijver’ ook de woorden ‘leven’ en ‘man’, ‘vrouw’ en ‘mensen’ relatief veel voorkomen in de recensies uit het *Algemeen Dagblad* in 1955.



Afb. 4: voorbeeld van een woordwolk

EXTRA OPTIES

Hierboven werd duidelijk dat door het corpus in te delen in mappen en sets (jaren, kranten, jaargangen van kranten), het mogelijk is om kleinere delen ervan te analyseren. Naast deze opties zijn er nog meer mogelijkheden om de recensies in kleinere groepen te analyseren. Zo kunnen per document verschillende extra variabelen genoteerd worden. Het is bijvoorbeeld mogelijk om voor elke recensie aan te geven welk genre boek besproken wordt. Vervolgens kunnen de recensies per genre eenvoudig geselecteerd worden om uitgebreider bekeken te worden. Ook kunnen verschillende genres zo met elkaar vergeleken worden. Een andere vergelijking waar we in geïnteresseerd waren is een eventueel verschil tussen recensies van verhalenbundels en recensies van romans. Door een variabele ‘roman of verhalen’ aan te maken, kunnen deze gegevens eenvoudig bekeken worden.

Naast het invoeren van extra variabelen is het ook mogelijk een memo met extra informatie te maken bij een document. Deze informatie kan vervolgens worden doorzocht. Door de naam van de criticus in een memo bij de recensie te zetten, kunnen alle recensies van een criticus eenvoudig door één zoekopdracht gevonden worden. Overigens geldt de mogelijkheid om memo's toe te voegen, en later te doorzoeken, niet alleen voor hele documenten, maar ook voor bijvoorbeeld delen ervan en voor codes.

MAXqda (en vergelijkbare software) biedt naast deze mogelijkheden om het corpus in te delen in kleinere sets (die vervolgens geanalyseerd kunnen worden), ook nog andere opties die relevant kunnen zijn voor, onder andere, tijdschriftonderzoek. Zo biedt het een aantal (beperkte) mogelijkheden om de coderingen van twee codeurs met elkaar te vergelijken en verschillende opties om de verdelingen van codes over een tekst visueel weer te geven. Door deze mogelijkheden is dergelijke software breed in te zetten, zowel voor kwalitatief onderzoek als voor (deels) kwantitatief onderzoek.