

# Hoe literair is het idee: meten is weten?

---

Steeds meer literatuurlijefhebbers en -uitgevers houden zich bezig met big data in het boekenvak. Wat betekent dit voor de schrijvers, uitgevers en lezers? Gaat de romantiek van de schrijver met kaars op een zolderkamer daarmee verloren?

---

In 1985 houdt de Amerikaanse schrijver Kurt Vonnegut een lezing, *The shape of stories*, waarin hij vertelt hoe literaire verhalen zijn opgebouwd. Eigenlijk is literatuur schrijven heel simpel, legt hij uit: je hebt drie plotlijnen waaruit je kunt kiezen. Op een krijtbord tekent hij twee strepen: een verticale lijn geeft de geluksfactor aan, de horizontale lijn geeft het tijdsverloop in het verhaal aan. De eerste vertelvariant is ‘Man in hole’: een personage is gelukkig, komt in problemen en klimt er weer uit. De tweede variant is ‘Boy meets girl’: een jongen staat in het midden van de geluksfactor, is enorm gelukkig wanneer hij verliefd wordt, er komen problemen dus hij wordt diep ongelukkig, als die voorbij zijn is hij gelukkiger dan hij aan het begin was. De derde variant is ‘Assepoester’: een meisje is diep ongelukkig, ze mag naar een feest dus stijgt de gelukslijn, bereikt een hoogtepunt in geluk wanneer ze met de prins danst, als dat opeens voorbij is, is ze weer zeer ongelukkig, maar door de herinnering aan het feest is ze gelukkiger dan eerst, dan past ze de schoen en bereikt het toppunt van geluk want ze trouwt met de prins. De laatste variant heeft duidelijk het grilligste geluksverloop.

### Vijf verhaallijnen

Toen de lezing werd gebundeld in *Man without a country* (2005) voegde Vonnegut er nog twee varianten aan toe: de ‘Kafka-Metamorfose variant’ – iemand is diep ongelukkig, wordt nog ongelukkiger en gaat dan dood, en de ‘Hamlet-variant’ – iemand is een prins, hij vermoordt iemand, wordt niet gestraft en sterft zelf, zonder dat je weet wat er intussen met hem gebeurt.

Vonnegut was zijn tijd vooruit. Inmiddels hebben computers op basis van zo’n 2000 Engelse dan wel in het Engels vertaalde werken geanalyseerd welke verhaallijnen er zijn. Aan de hand van bepaalde woorden waaruit de emotionele lading van een tekst blijkt, of in ieder geval de emotionele ontwikkeling die het personage ondergaat, blijkt dat er vijf verhaalstructuren zijn: Van arm naar rijk (erbovenop komen), van rijk naar arm (neergang), Man in a hole (val en er dan uit komen), Icarus (stijgen en dan ten val komen), Assepoester (stijgen, vallen en stijgen), Oedipus (vallen, stijgen, ten val komen) (LaFrance 2016).

Stel dat elke auteur vanuit deze vijf schema’s moet werken: je zou om minder een writer’s block

krijgen. De varianten laten echter wel zien welke literaire conventies we erop nahouden.

### Algoritmen analyseren stijl

Er zijn talloze mogelijkheden wanneer je werkt met big data, legt Karina van Dalen-Oskam uit. Zij leidt bij het Huygens Instituut voor Nederlandse Geschiedenis (Huygens ING) het project *The riddle of literary quality*, waarin met behulp van algoritmen teksten worden geanalyseerd. ‘Uitgevers kunnen met algoritmen een stijl analyseren, kijken waar iets op lijkt en of het past binnen je fondslijn bijvoorbeeld.’ Uitgevers kunnen zo eenvoudiger inschatten of een roman in een niche komt of niet. Een ander voordeel ziet ze in de beoordeling van een manuscript. Vele ongevraagde manuscripten die nu bij een uitgeverij binnenkomen, worden niet gelezen of na een pagina of tien al weggelegd. Van Dalen-Oskam: ‘Nu maakt een manuscript tenminste nog een kans om door een computer gelezen te worden.’ Een van de onderzoeken die momenteel bij het Huygens ING worden gedaan, is kijken wat literaire boeken wel hebben dat niet-literaire boeken ontberen. ‘Je kunt de canon koppelen aan de genre aanduiding van nu en bijvoorbeeld teksten van Couperus meten om vast te stellen in hoeverre die afwijken van moderne chicklit of literaire fictie’, legt Van Dalen-Oskam uit. Met algoritmen kun je onderzoeken welke genre aanduidingen worden toegepast bij verschillende auteurs en welk type boek sneller als literair wordt aangemerkt of juist niet, maar aan de hand van de woordfrequentie ook kijken welk soort boek al dan niet literair wordt gevonden. Onderzoek naar het woordgebruik kan tonen welke auteur boeken onder pseudoniem heeft geschreven. Elke auteur heeft namelijk een bepaald woordgebruik, zelfs aan de hand van lidwoorden en voegwoorden valt volgens Van Dalen-Oskam de auteur te herkennen. Zo werd ontdekt dat J.K. Rowling achter het pseudoniem Robert Galbraith zat en wie de boeken van Elena Ferrante heeft geschreven. In beide gevallen ging er iets verloren aan de mystiek rond de schrijver. Maar is dat erg? Is het idee van een machine die literatuur ontrafelt of zelfs schrijft eng, of eenvoudigweg de toekomst, of misschien niets meer dan een handig instrument waar schrijvers, uitgevers en lezers hun voordeel mee kunnen doen? →

### Experimentele resultaten

In Nederland gingen Ronald Giphart en Arnon Grunberg hierin vooralsnog het verst. Giphart schreef in samenwerking met de Universiteit Antwerpen en het Meertens Instituut en gebruikmakend van de robot Asibot (een verwijzing naar sciencefictionschrijver Isaac Asimov), een tiende verhaal voor Asimovs verhalenbundel *Ik robot*. Het ging duidelijk niet vanzelf: ‘Ik heb geprobeerd Asibot zoveel mogelijk zelf te laten schrijven, maar het moest natuurlijk wel een verhaal worden’, aldus Giphart (Lindeboom 2018). Waar het de computer aan ontbrak was verbanden leggen om een logisch plot te maken, dat hangt samen met verbeelding.

Grunbergs novelle *Het bestand* was aanleiding om ook de ervaring van de lezer te onderzoeken: na verschijning deed de Universiteit van Amsterdam aan de hand van hersenactiviteit, hartritme en vingervochtigheid onderzoek naar de emoties bij de lezer. Het meetresultaat was weinig verrassend, zoals bij de meeste boeken van Grunberg was de lezer niet blij na het lezen van de roman en scènes met erg veel geweld riepen meer emoties op dan een scène waarin een omgeving werd beschreven. Doordat Grunberg zich meer richtte op de emotie van de lezer dan op het verhaal zelf, was het resultaat wel dat dit niet bepaald Grunbergs kwalitatief sterkste boek was.

In beide gevallen leverde het meten, weten en robotiseren weinig meer op dan wat experimentele resultaten. Mocht de schrijver een machine worden, dan is de lezer toch vooral gebaat bij die van Italo Calvino. Hij voorspelde een literatuurmachine die genoeg zou krijgen van de gang-

bare vormen en zou inspelen op de menselijke behoefte: *disorder*. ‘De ware literaire machine zal in staat zijn om uit zichzelf wanorde te creëren, als reactie op de orde die eraan voorafging: een machine dus die avant-gardewerk produceert wanneer het circuit vastloopt vanwege een teveel aan classicisme’ (Calvino 1986).

### Aantal worstsellers verminderen

Uitgevers zouden niet veel aan zo’n machine hebben. Patrick Swart, directeur van uitgeefconcern WPG, waarin onder andere De Bezige Bij en A.W. Bruna zijn ondergebracht, gaf in een interview aan dat algoritmen mogelijkerwijs redacteurs kunnen ondersteunen. ‘We bekijken bijvoorbeeld hoeveel geweld erin zit, hoeveel seks, of het een happy ending heeft, waar het verhaal zich afspeelt en hoe divers het woordgebruik is. Je kunt hier eindeloos op doorgaan. Je kunt er recensies aan toevoegen, de verkoop aantallen, enzovoorts. Als je maar genoeg informatie toevoegt, krijg je met behulp van algoritmes een heel interessant inzicht in de verschillen en overeenkomsten tussen boeken. Zo kun je bepalen wat een boek uniek en onderscheidend maakt. Als je straks een manuscript in de machine stopt, kan de computer de recensie schrijven’ (Chin et al. 2017). Dit is volgens hem niet de manier om bestsellers te creëren, maar wel om het uitgeven van *worstellers* te verminderen.

Ook het andere grote uitgeefconcern VBK (met onder andere Atlas Contact en Ambol Anthos) is met big data bezig. Voormalig algemeen directeur van VBK Wiet de Bruijn stelde: ‘We onderzoeken aan de hand van big data wat lezers prettig vinden om in bepaalde genres te lezen. Daarmee kunnen we schrijvers tips geven: zorg wel dat op pagina 10 dit of dat gebeurd is, anders haken lezers misschien af. Waarschijnlijk werkt het alleen voor genres die vaak volgens vrij vaststaande patronen werken, zoals thrillers en fantasy’ (Chin et al. 2017).

Na deze uitspraken werd er door menigeen gegniffeld – vooral het woord *worstseller* was op social media een dankbaar mikpunt van woordspelingen – en er werd gewezen op het gevaar van (excusez le mot) eenheidsworst. Die vrees is eigenlijk ongegrond, meent Van Dalen-Oskam. Je kunt als uitgever aan de hand van de geleverde data juist op zoek gaan naar de afwijkingen en daar je uitgeefbeleid op aanpassen.

**‘Als je straks een manuscript in de machine stopt, kan de computer de recensie schrijven’**

De onderzoeken bieden bovendien veel mogelijkheden. Corina Koolen doet bijvoorbeeld in haar proefschrift *Reading beyond the female* sociologisch onderzoek en zet overtuigend neer hoe genres gekoppeld worden aan man/vrouw als auteur en hoe de literaire conventie nog steeds is dat mannen kwalitatief betere romans schrijven. Zowel lezers als uitgevers spelen hierin een rol. Ze geven mannelijke auteurs eerder uit als literatuur en brengen romans van vrouwen vaker op de markt als romantische fictie. Uit de computeranalyses die Koolen deed, bleek namelijk dat de stijl wordt bepaald door genre, dialoog en narratief en niet door het gegeven of een man dan wel vrouw de tekst had geschreven. En toch worden werken van mannen als literair hoogstaander beoordeeld: 'Het gaat vooral om perceptie: vrouwelijke auteurs zijn geen andere soort, zij schrijven – net als mannelijke auteurs – in de stijl van het genre dat zij beoefenen' (Koolen 2018). De uitgever kan zijn of haar voordeel met deze kennis doen, mits je als uitgever vindt dat je een rol hebt in genderdiscussies.

### Tij keren

Met de data-analyses tot dusver had de computer het volgens Van Dalen-Oskam in 80 procent juist wanneer het ging om commerciële potentie. Het is niet zo dat uitgevers hier hun oren naar moeten laten hangen, maar met die gegevens kun je achteraf bekijken wat er misging in bijvoorbeeld de marketing van een boek en daar lessen uit trekken.

De vraag blijft wat de lezer eraan heeft en wat het nut is om te weten hoe lezen werkt in je hoofd. Bij het onderzoek van Koolen is het vooral confronterend voor de lezer dat hij/zij een literaire seksist blijkt, maar dankzij die kennis kun je het tij ook proberen te keren.

Pijnlijk is het voor schrijvers wanneer ze rekening moeten gaan houden met de nukken van de lezer, zoals De Bruijn suggereerde, omdat lezers anders misschien afhaken. Dat afhaken is immers meetbaar. In 2014 maakte Kobo, een abonnementsdienst voor e-boeken, bekend dat van alle Britse kopers van *Het puttertje* van Donna Tartt slechts 44 procent het boek daadwerkelijk had uitgelezen (Flood 2014). Voor auteurs is dat vervelend, voor uitgevers niet. Umberto Eco staat bekend als de bestverkochte niet gelezen auteur, en liet zich er ook op voorstaan dat hij zijn publiek wilde uittesten voor hij ze spanning gaf.

Maar je moet er niet aan denken dat je er als schrijver rekening mee moet houden. Net als die vijf vaste plotlijnen kan zo iets alleen maar een writer's block in de hand werken. En uitgevers? Het is te hopen dat die juist op zoek gaan naar de niche, en wie weet worden er dan vanzelf meer verhaallijnen aan de conventies toegevoegd. ●

### Literatuur

- Calvino, I. (1986) 'Cybernetics and Ghosts'. In: *The uses of literature*, 3-27.
- Chin, H. en T. Jaeger (2017) 'Big data en algoritmes moeten worstsellers gaan voorkomen'. In: *NRC Handelsblad*, 23 maart.
- Flood, A. (2014) 'Ebooks can tell which novels you didn't finish'. In: *The Guardian*, 10 december.
- Grunberg, A. (2015) *Het bestand*. Amsterdam: Nijgh & Van Ditmar.
- Koolen, C. (2018) *The riddle of literary quality*. Den Haag: Huygens Instituut ING.
- LaFrance, A. (2016) 'The six main arcs in storytelling, as identified by an A.I.'. In: *The Atlantic*, 12 juli.
- Lindeboom, M. (2018) 'Ronald Giphart over robots, uitgevers en zijn nieuwe boek'. Op: *Hebban*, 4 april. ([www.hebban.nl/artikelen/ronald-giphart-over-robots-uitgevers-en-zijn-nieuwe-boek](http://www.hebban.nl/artikelen/ronald-giphart-over-robots-uitgevers-en-zijn-nieuwe-boek))
- Vonnegut, K. (2005) 'Here is a lesson in creative writing'. In: *A man without a country*, 23-38.