

# Verkiezingsvocabulaire vergeleken

De taal van de verkiezingsprogramma's voor de Tweede Kamerverkiezingen van 17 maart 2021  
28 februari 2021

Jan van Casteren

## 1 Introductie

Op woensdag 17 maart 2021 kiest Nederland de leden van de Tweede Kamer.<sup>1</sup> In de aanloop hier-naartoe presenteren de deelnemende politieke partijen een verkiezingsprogramma, waarin de plannen voor de komende periode worden uiteengezet. Middels dit artikel presenteer ik een kwantitatieve analyse van de overeenkomsten en verschillen tussen de taal die door de verscheidene partijen wordt gehanteerd.<sup>2</sup> Hiervoor gebruik ik technieken uit het domein *Text Mining*: het automatisch extraheren van informatie uit grote hoeveelheden tekst [Hotho et al., 2005].

Het doel van dit onderzoek is om een beeld te krijgen van de onderlinge verhoudingen tussen de verschillende politieke partijen. De gebruikte taal in de desbetreffende verkiezingsprogramma's dient als maatstaf voor deze vergelijking. Om de reikwijdte van het onderzoek redelijk te houden, zijn alleen de politieke partijen die in 2017 ten minste een zetel hebben bemachtigd binnen beschouwing genomen.

De drie vragen die ik tracht te beantwoorden luiden als volgt:

1. Welke partijen hanteren vergelijkbare termen?
2. In hoeverre is de taal die een partij hanteert veranderd ten opzichte van de vorige verkiezingscampagne?
3. In hoeverre zijn de overeenkomsten in taal tussen partijen veranderd ten opzichte van de vorige campagne?

Deze vragen worden beantwoord op de volgende wijze: in Sectie 2 beschrijf ik welke methoden zijn gebruikt om bovengenoemde vragen te beantwoorden. Ik presenteer en bespreek de resultaten in Sectie 3. In Sectie 4 wordt de zeggingskracht van het onderzoek bediscussieerd. Tevens geef ik hier aan hoe een eventueel vervolgonderzoek uitgevoerd zou kunnen worden met een toegenomen nauwkeurigheid. Tot slot voorzie ik in Sectie 5 in een samenvatting van de verzamelde antwoorden op bovengenoemde onderzoeksvragen.

<sup>1</sup> Dit jaar zal het tevens mogelijk zijn de voorafgaande dagen een stem uit te brengen.

<sup>2</sup> Alle relevante code is gepubliceerd op <http://github.com/vancasteren/verkiezingsvocab>

## 2 Methoden

In deze sectie bespreek ik hoe ik te werk ben gegaan. Hierin te onderscheiden zijn respectievelijk het voorbereidend en het analytisch werk.

Het voorbereidend werk besloeg de volgende stappen: ten eerste heb ik van elke partij, die in 2017 in de Tweede Kamer verkozen is, de verkiezingsprogramma's verzameld. Ten tweede is uit elk document de tekst geëxtraheerd. Omdat onze interesse uitgaat naar de lopende tekst, heb ik de kop- en voetteksten van iedere tekst weggefilterd. Ook de namen van de partijen zelf zijn uit de lopende tekst verwijderd.

Als derde stap zijn uit elke tekst alle woorden *gelemmatiseerd*<sup>3</sup> en geteld. De lemma's zijn bepaald met behulp van Stanza. [Qi et al., 2020]. De precieze nauwkeurigheid van Stanza's lemmatizer voor het Nederlands is helaas niet bekend. Een behoudende schatting zou zijn dat de lemmatizer voor 90% accuraat is. Dat is op zich niet onoverkomelijk, maar wel belangrijk om te vermelden.

Tot slot zijn alle zgn. stopwoorden uit de lijst met lemma's verwijderd. Stopwoorden zijn termen die zeer regelmatig in teksten voorkomen, maar nauwelijks betekenis in zich dragen.<sup>4</sup> Het verwijderen van de stopwoorden zorgt ervoor dat de teksten op een betekenisvolle manier kunnen worden vergeleken.

De getelde en gefilterde lemma's per verkiezingsprogramma vormen de data van dit onderzoek. De verzameling van de data is het sluitstuk van de voorbereidende fase. Om de data te kunnen interpreteren, zijn de volgende stappen ondernomen:<sup>5</sup>

0. Per tekst zijn een aantal basisstatistieken verzameld: het aantal woorden, het aantal unieke lemma's en de gemiddelde frequentie van de lemma's.
1. Op basis van de lemmafrequenties is de overeenkomstigheid tussen de teksten bepaald aan de hand van de onderlinge cosinusgelijkenis. Dit is een gebruikelijke methode om de gelijkenis tussen documenten te bepalen [Gong et al., 2018].

<sup>3</sup> Dat wil zeggen: de woorden zijn teruggebracht hun lemma, het 'woordenboekwoord' van het woord in kwestie. 'Zegt' wordt dus 'zeggen', 'loop' wordt 'lopen', 'bomen' wordt 'boom' en 'doen' blijft 'doen'.

<sup>4</sup> <https://github.com/stopwords-iso/stopwords-nl>

<sup>5</sup> Bij wijze van beroepsdeformatie, en om de punten 1 t/m 3 overeen te laten komen met de bijbehorende onderzoeksvragen, begint deze telling bij 0.

Partij	Aantal woorden	Unieke woorden	Gem. overeenkomst	Consistentie	Verschuiving (med.)
CU	31.479	6.945	81,8%	93,2%	+3,3 pp
D66	30.529	6.404	81,7%	93,3%	+3,0 pp
PvdA	20.948	6.127	80,6%	89,3%	+2,2 pp
GL	12.244	3.654	79,2%	87,0%	+4,3 pp
VVD	19.073	4.886	79,2%	88,1%	+1,2 pp
SGP	28.963	6.810	78,6%	94,3%	+2,3 pp
CDA	14.292	3.920	74,8%	87,1%	-1,0 pp
DENK	6.518	2.520	71,8%	73,1%	+7,0 pp
PvdD	20.081	5.529	71,5%	92,6%	+4,4 pp
SP	4.382	1.737	71,1%	84,8%	-3,5 pp
FVD	7.617	2.996	67,8%	79,8%	+1,8 pp
50PLUS	1.890	992	67,4%	70,6%	+8,4 pp
PVV	3.992	1.779	66,4%	44,3%	+40,8 pp
Gem.	15.539,1	4.176,8	74,8%	82,9%	+3,9 pp (med.)

Tab. 1: Overzicht van de statistieken per partij(programma)

Een gelijkenis van 0% betekent dat er in het geheel geen gelijkenis bestaat tussen de vocabulaires van twee teksten; een gelijkenis van 100% houdt in dat de teksten precies even vaak precies dezelfde woorden bevatten.<sup>6</sup> Partijen met een hoge cosinusgelijkenis hebben dus veel taal gemeen, partijen met een lage weging weinig.

2. De cosinusgelijkenis is daarnaast ook berekend tussen het huidige en vorige verkiezingsprogramma van elke partij. Welke partij houdt woord, wie beweegt mee met de tijd?
3. De in stap 0 en 1 genoemde handelingen zijn eveneens uitgevoerd voor de verkiezingsprogramma's voor de vorige Tweede Kamerverkiezingen in 2017. Hiermee hoop ik een licht te werpen op tussentijdse verschuivingen in het politieke landschap.

### 3 Resultaten

**Basisstatistieken** De basisstatistieken per partij-programma (aantal woorden, aantal unieke woorden) zijn zichtbaar in Tabel 1. Ook de gemiddelde overeenkomst met andere partijen, consistentie (de cosinusgelijkenis met het eigen vorige verkiezingsprogramma) en de mediaan van de verschuiving per partij zijn hier toegevoegd. In de onderste rij zijn voor elk van de genoemde waardes de gemiddelden opgenomen. Omdat er wat de verschuivingen betreft een aantal uitschieters waren, heb ik voor de desbetreffende kolom in plaats van het gemiddelde de mediaan gebruikt.<sup>7</sup>

Ik bespreek eerst het aantal woorden en aantal unieke woorden per partij. De overige drie kolommen worden nader toegelicht in de volgende paragrafen.

<sup>6</sup> Het is gebruikelijk cosinusgelijkenis weer te geven als een waarde tussen 0 en 1. Omwille van toegankelijkheid heb ik ervoor gekozen om percentages te gebruiken.

<sup>7</sup> De verschuivingen ten opzichte van 2017 bevatten uitschieters zoals die verder niet voorkomen. Daarom gebruik ik alleen voor deze categorie de mediaan. Dit wordt nader toegelicht in de paragraaf 'Verschuivingen van de overeenkomsten'.

Aan het aantal woorden per partij valt met name op dat er grote verschillen zijn wat de breedsprakigheid van de verschillende partijen betreft. Voor de volledigheid is ook de gemiddelde overeenkomst met de overige partijprogramma's hier vermeld, evenals de mate van consistentie t.o.v. 2017. Deze twee statistieken worden verderop apart besproken.

Geen van de partijen die minder dan 10.000 woorden gebruiken heeft ooit ministers geleverd aan een kabinet. Van deze partijen heeft alleen de PVV deelgenomen aan een coalitie, als gedoogpartner weliswaar. Van alle deelnemende partijen die ooit aan een kabinet hebben deelgenomen (CDA, CU, D66, PvdA, VVD), gebruikt het CDA de minste woorden.

De ChristenUnie heeft het rijkste vocabulaire; deze partij gebruikt het grootste aantal verschillende termen. 50PLUS gebruikt het minste aantal unieke termen en is tevens het kortst van stof.

**Onderlinge cosinusgelijkenis** De onderlinge cosinusgelijkenissen tussen de partijen zijn zichtbaar in Figuur 1. De gemiddelde gelijkenis bedraagt 74,77%,  $\sigma = 8,61\%$ .

De programma's van CU en D66 komen het meest overeen, met een gelijkenis van 92,3%. De grootste gelijkenis die hierop volgt, is die tussen CU en SGP (91,5%).

De grootste kloof is te vinden tussen 50PLUS en de PVV, met een gelijkenis van slechts 58,4%. Tussen 50PLUS en FVD botert het evenmin (58,8%).

Verder is het bij de PvdD interessant dat deze partij in ongeveer gelijke mate verwantschap toont met CU, D66, GL en PvdA. Ook de hoge mate van overeenkomst tussen D66 en het CDA is opvallend (86,0%). Eveneens merkwaardig is dat de SGP een hogere gelijkenis erop nahoudt met de PvdA en D66, dan met het CDA. Tevens wens ik op te merken dat allevier de huidige coalitiepartijen (CDA, CU, D66, VVD) meer dan 80% gelijkenis met elkaar vertonen.

De ChristenUnie is voor liefst zes partijen de meest nabij gelegen partij (50PLUS, CDA, D66, DENK, PvdD, SGP). D66 ligt voor vier partijen het dichtste

50PLUS	100.0%	66.3%	75.1%	72.7%	65.3%	58.8%	68.0%	72.3%	66.3%	58.4%	72.1%	63.9%	69.4%
CDA	66.3%	100.0%	86.5%	86.0%	69.9%	66.7%	80.7%	81.8%	67.7%	61.6%	81.2%	69.1%	80.7%
CU	75.1%	86.5%	100.0%	92.5%	77.6%	71.6%	87.4%	90.2%	78.6%	66.9%	91.7%	76.4%	87.7%
D66	72.7%	86.0%	92.5%	100.0%	76.6%	72.5%	90.1%	90.6%	78.4%	67.4%	88.3%	76.3%	89.0%
DENK	65.3%	69.9%	77.6%	76.6%	100.0%	67.3%	75.7%	77.3%	66.1%	68.2%	73.8%	67.8%	76.0%
FVD	58.8%	66.7%	71.6%	72.5%	67.3%	100.0%	68.6%	70.2%	62.7%	69.7%	71.6%	60.2%	73.9%
GL	68.0%	80.7%	87.4%	90.1%	75.7%	68.6%	100.0%	89.5%	78.0%	67.1%	81.0%	81.4%	83.3%
PvdA	72.3%	81.8%	90.2%	90.6%	77.3%	70.2%	89.5%	100.0%	77.8%	68.4%	84.4%	81.6%	83.4%
PvdD	66.3%	67.7%	78.6%	78.4%	66.1%	62.7%	78.0%	77.8%	100.0%	64.1%	74.7%	68.0%	75.2%
PVV	58.4%	61.6%	66.9%	67.4%	68.2%	69.7%	67.1%	68.4%	64.1%	100.0%	67.9%	64.2%	73.2%
SGP	72.1%	81.2%	91.7%	88.3%	73.8%	71.6%	81.0%	84.4%	74.7%	67.9%	100.0%	71.3%	85.3%
SP	63.9%	69.1%	76.4%	76.3%	67.8%	60.2%	81.4%	81.6%	68.0%	64.2%	71.3%	100.0%	73.1%
VVD	69.4%	80.7%	87.7%	89.0%	76.0%	73.9%	83.3%	83.4%	75.2%	73.2%	85.3%	73.1%	100.0%
	50PLUS	CDA	CU	D66	DENK	FVD	GL	PvdA	PvdD	PVV	SGP	SP	VVD

Fig. 1: Onderlinge cosinusgelijkenis tussen verkiezingsprogramma's

bij (CU, GL, PvdA, VVD), VVD voor twee (FVD en PVV) en PvdA voor een (SP).

#### Gemiddelde overeenkomst met andere partijen

Figuur 2 toont voor elke partij de gemiddelde overeenkomst met de andere partijen in 2017 en 2021. Deze waarden zijn de gewogen gemiddelden van de onderlinge cosinusgelijkenissen van elke partij in vergelijking tot de andere partijen. Deze gemiddelden zijn tevens opgenomen in Tabel 1, onder de noemer 'Gem. overeenkomst'.

Zes partijen vormen tezamen het midden. CU,

D66, GL, PvdA, SGP en VVD hebben gemiddeld 80% van hun vocabulaire gemeen met de overige politieke partijen. Een tweede zestal, bestaande uit 50PLUS, DENK, FVD, PvdD, PVV en SP, scoort rond de 70%. Het CDA bevindt zich in het midden tussen deze twee groepen. Opvallend is dat alle partijen met een lagere gemiddelde overeenkomst dan het CDA, relatief nieuwe partijen zijn (opgericht na 2000), met uitzondering van SP. De PVV heeft het meest unieke taalgebruik.

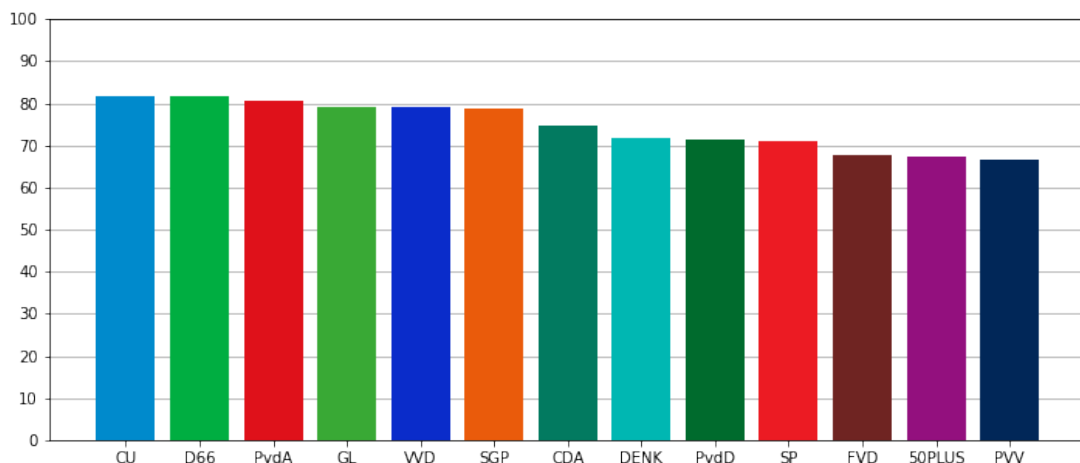


Fig. 2: Gemiddelde overeenkomst met overige verkiezingsprogramma's (cosinusgelijkenis)

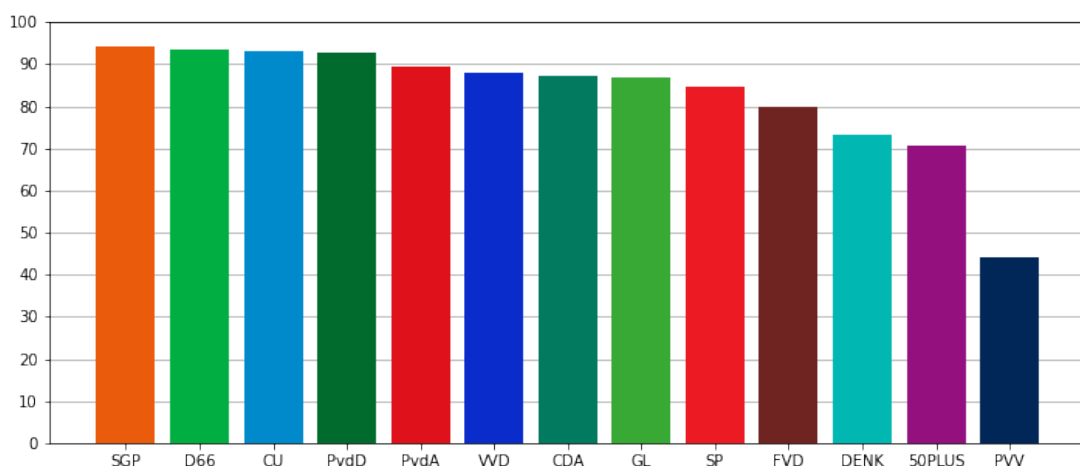


Fig. 3: Cosinusgelijkenis per partij ten opzichte van het vorige eigen verkiezingsprogramma

**Consistentie** In Figuur 3 wordt weergegeven in hoeverre het verkiezingsprogramma van een partij overeenkomt met het vorige programma. Deze waarden zijn eveneens zichtbaar in Tabel 1. Wat de consistentie van het vocabulaire betreft is de SGP koploper, gevolgd door D66 en daarna ChristenUnie. De klaarblijkelijke inconsistente van de PVV is vooral te wijten aan de vervijftigvoudiging in lengte van diens verkiezingsprogramma t.o.v. 2017. Niettemin valt het op dat het gros van de partijen die na 2000 zijn opgericht nog op zoek lijkt naar vastigheid in de formulering. De Partij voor de Dieren is hierop de uitzondering.

**Verschuivingen van de overeenkomsten** In Figuur 4 is te zien hoe de onderlinge verhoudingen zijn veranderd ten opzichte van de vorige verkiezingen. Hier is te zien hoe de PVV nu veel meer taal gemeen heeft met de andere partijen, zoals eerder bespro-

ken. Dit betekent echter ook dat, vanuit de overige partijen bezien, een toenemende overeenkomst met de PVV niet al te betekenisvol geacht moet worden. Wel is het interessant te benoemen dat de grootste toename qua overeenkomst voor de PVV die met de VVD betreft (+46,0pp), gevolgd door die met de PvdA (+44,5pp).

De PVV daargelaten heeft de grootste mate van toenadering plaatsgevonden tussen 50PLUS en DENK (+13,8pp). Ook tussen 50PLUS en de PvdD lijkt er iets te zijn gaan bloeien (+11,7pp), evenals tussen DENK en Forum voor Democratie (+10,1pp). De grootste mate van verwijdering heeft plaatsgevonden tussen de SP en de SGP (−8,4pp), gevolgd door SP en CDA (−7,1pp). Wat de taal betreft, noteert de SP sowieso het grootste aantal negatieve verschuivingen; alleen ten opzichte van 50PLUS, DENK en GL heeft deze partij enige toenadering gevonden.

Bij het CDA is het opmerkelijk dat de overeen-

50PLUS	0.0pp	2.2pp	10.0pp	9.1pp	13.8pp	2.3pp	8.0pp	8.8pp	11.7pp	36.9pp	7.6pp	5.3pp	6.8pp
CDA	2.2pp	0.0pp	-2.0pp	0.0pp	-3.4pp	-0.1pp	4.2pp	-4.3pp	2.3pp	38.1pp	-1.9pp	-7.1pp	-3.5pp
CU	10.0pp	-2.0pp	0.0pp	3.0pp	6.4pp	2.5pp	4.4pp	0.9pp	5.5pp	43.3pp	3.6pp	-4.5pp	1.6pp
D66	9.1pp	0.0pp	3.0pp	0.0pp	5.9pp	1.4pp	5.9pp	2.6pp	4.7pp	42.0pp	2.9pp	-3.8pp	-1.3pp
DENK	13.8pp	-3.4pp	6.4pp	5.9pp	0.0pp	10.1pp	9.0pp	5.1pp	6.6pp	39.6pp	9.1pp	7.0pp	6.9pp
FVD	2.3pp	-0.1pp	2.5pp	1.4pp	10.1pp	0.0pp	3.9pp	1.3pp	3.0pp	41.8pp	-0.6pp	-4.9pp	0.8pp
GL	8.0pp	4.2pp	4.4pp	5.9pp	9.0pp	3.9pp	0.0pp	5.8pp	1.2pp	40.3pp	2.0pp	0.7pp	3.9pp
PvdA	8.8pp	-4.3pp	0.9pp	2.6pp	5.1pp	1.3pp	5.8pp	0.0pp	5.7pp	44.5pp	1.7pp	-1.3pp	-1.0pp
PvdD	11.7pp	2.3pp	5.5pp	4.7pp	6.6pp	3.0pp	1.2pp	5.7pp	0.0pp	39.3pp	2.6pp	-3.2pp	4.2pp
PVV	36.9pp	38.1pp	43.3pp	42.0pp	39.6pp	41.8pp	40.3pp	44.5pp	39.3pp	0.0pp	41.4pp	39.2pp	46.0pp
SGP	7.6pp	-1.9pp	3.6pp	2.9pp	9.1pp	-0.6pp	2.0pp	1.7pp	2.6pp	41.4pp	0.0pp	-8.4pp	-1.7pp
SP	5.3pp	-7.1pp	-4.5pp	-3.8pp	7.0pp	-4.9pp	0.7pp	-1.3pp	-3.2pp	39.2pp	-8.4pp	0.0pp	-6.4pp
VVD	6.8pp	-3.5pp	1.6pp	-1.3pp	6.9pp	0.8pp	3.9pp	-1.0pp	4.2pp	46.0pp	-1.7pp	-6.4pp	0.0pp
	50PLUS	CDA	CU	D66	DENK	FVD	GL	PvdA	PvdD	PVV	SGP	SP	VVD

Fig. 4: Veranderingen in cosinusgelijkenis t.o.v. 2017

komst met GL is toegenomen (+4,2pp), waar die met de PvdA juist is afgenomen (-4,3pp). Het CDA is geen van haar coalitiepartners genaderd; de overeenkomst met CU en VVD is afgenomen (-2,0pp resp. -3,5pp), die met D66 gelijk gebleven (0,0pp).

Verder valt het op dat de VVD haar imago als “GroenRechts” terug heeft opgevijseld, blijkens een toegenomen overeenkomstigheid met PvdD en GroenLinks (+4,2pp resp. +3,9pp). De PvdD is op haar beurt opgeschoven naar het midden; alleen voor de SP noteert deze partij een afgenomen overeenkomstigheid. Deze tendens is nog sterker zichtbaar

bij 50PLUS; deze partij ligt nu dichterbij alle andere partijen. Omdat 50PLUS relatief veel aan het eigen programma heeft veranderd<sup>8</sup> ligt het voor de hand dat deze beweging vooral een oorsprong heeft binnen 50PLUS zelf. Ook GL kent louter toegenomen overeenkomsten met de overige partijen. DENK ligt nu, net als de PvdD, dichterbij alle partijen behalve één; in het geval van DENK betreft dit het CDA (-3,4pp). De ChristenUnie is alleen van CDA en van de SP (-2,2pp resp. -4,2pp) verwijderd ge-

<sup>8</sup> Dit wordt bijvoorbeeld duidelijk uit Figuur 3.

	2021		2017	
Partij	Partij	Gelijkenis	Partij	Gelijkenis
50PLUS	CU	74,9%	CU	65,1%
CDA	CU	86,3%	CU	88,5%
CU	D66	92,3%	D66	89,5%
D66	CU	92,3%	VVD	90,3%
DENK	CU	77,5%	CDA	73,3%
FVD	VVD	73,9%	VVD	73,1%
GL	D66	90,1%	D66	84,2%
PvdA	D66	90,6%	CU	89,3%
PvdD	D66	78,4%	GL	76,8%
PVV	VVD	73,2%	DENK	28,6%
SGP	CU	91,5%	CU	88,1%
SP	PvdA	81,6%	PvdA	82,9%
VVD	D66	89,0%	D66	90,3%

Tab. 2: Grootste gelijkenis per partij in 2021 en 2017.

raakt.

Om de gemiddelde verschuiving per partij te berekenen heb ik niet gewerkt met het gewogen gemiddelde, maar met de mediaan. Deze keuze is gebaseerd op het gegeven dat één van de dertien partijen nogal is uitgeschoten wat de verschuivingen betreft; dit is een gevolg van de keuze van de PVV om in 2021 meer dan honderd keer zoveel woorden te gebruiken als in 2017. De mediale verschuiving per partij is opgenomen in Tabel 1.

Gemiddeld genomen zijn alleen de SP en het CDA wegbewogen van de overige partijen (−3,5pp resp. −1,0pp). De overige partijen zijn juist overwegend naar elkaar toe bewogen. Naast de PVV zijn DENK en 50PLUS op dit gebied de uitschieters; de mediale verschuiving voor DENK is +7,0pp, voor 50PLUS is deze zelfs +8,4pp.

**Grootste gelijkenis per partij; 2021 & 2017** In Tab. 2 staat apart voor iedere partij genoteerd welke partij het dichtste bij ligt, en met welke mate van gelijkenis. In deze tabel is te zien hoe CU en D66 in 2021 hun posities in het midden hebben verstevigd ten opzichte van 2017. Zo was GroenLinks voor de PvdD de partij met de grootste mate van overeenkomst, waar D66 die rol nu heeft overgenomen; D66 is juist ‘geflipt’ van de VVD naar de CU.

## 4 Discussie

De grote variatie in breedsprakigheid tussen de verschillende partijen vormt een grote methodologische uitdaging voor een onderzoek zoals hier beschreven. Het langste partijprogramma beslaat meer dan tweehonderd bladzijden (D66); het kortste partijprogramma beslaat er vijftientig (50PLUS). Hoe meer bladzijden er zijn, hoe meer kans een auteur heeft om verschillende woorden te gebruiken. In het programma van D66 staan dus heel veel woorden die niet voorkomen in dat van 50PLUS. Het kan echter niet zomaar worden aangenomen dat deze woorden ook afwezig zouden zijn geweest in 50PLUS’ pro-

gramma, als de lieden van 50PLUS een programma hadden geschreven van dezelfde lengte als die van D66. De lengte van het document beïnvloedt dus hoe de frequentieverdeling van de woordenschat eruit ziet. In een vervolgonderzoek zou het interessant kunnen zijn meer aandacht te besteden aan methoden die lengteverschillen tussen teksten aan invloed doen inboeten. Het gebruik van *word embeddings* zou hier bijvoorbeeld aan kunnen bijdragen.<sup>9</sup>

Bij de verschuivingen van overeenkomst is het belangrijk te benadrukken dat het, langs de hier bewandelde weg, niet altijd met zekerheid na is te gaan welke van twee partijen verantwoordelijk is voor een toe- of afgenomen overeenkomst. Om dit uit te kunnen wijzen, zou nader onderzoek nodig zijn.

Verder dient opgemerkt te worden dat ik mij door kennis van het politieke landschap zo min mogelijk heb laten hinderen. Hier zie ik echter wel een mogelijkheid voor vervolgonderzoek. Het zou bijvoorbeeld interessant kunnen zijn om na te gaan in hoeverre er typisch ‘linkse’ dan wel ‘rechtse’ woorden bestaan. Voor een vervolgonderzoek dunkt dat me een boeiende onderneming.

Tot slot hecht ik eraan te benadrukken dat dit onderzoek het taalgebruik van de verschillende politieke partijen betreft, en dus niet de precieze standpunten. Een representatie van wat men daadwerkelijk van deze onderwerpen vindt, zult u in deze getallen niet tegenkomen. De cosinusgelijkenissen die hier worden gemeten representeren in hoeverre men aandacht besteedt aan dezelfde onderwerpen schrijft.

## 5 Conclusie

Middels dit onderzoek heb ik getracht een licht te schijnen op drie zaken, die te maken hebben met de taal die politieke partijen hanteren in hun verkiezingsprogramma. Ten eerste heb ik laten zien in hoeverre partijen hetzelfde of juist een verschillend vocabulaire hanteren. Vervolgens heb ik getoond in hoeverre de verschillende partijen in hun programma’s boodschapvast zijn. Tot slot heb verschuivingen in de onderlinge verhoudingen tussen de partijen inzichtelijk gemaakt. Ik zal de het meest in het oog springende observaties kort herhalen.

De ChristenUnie is qua taal de ultieme middenpartij van deze verkiezingen. Het verkiezingsprogramma van CU komt gemiddeld het meest overeen met die van de overige partijen; daarnaast is de CU voor liefst zes medepartijen het meest nabij gelegen. D66 is een goede tweede middenpartij; D66 heeft de op-één-na-hoogste gemiddelde overeenkomst en is tevens voor vier partijen het meest nabij gelegen. Het vocabulaire van de PVV is het meest eigen, gevolgd door resp. 50PLUS en FVD.

Relatief nieuwe partijen zijn over het algemeen minder boodschapvast: ten opzichte van 2017 hebben PVV, 50PLUS, DENK en FVD de grootste verandering in taal doorgemaakt. De SGP is juist een

<sup>9</sup> Zie bijvoorbeeld Kusner et al. [2015].

baken van vasthoudendheid, net als D66, CU en de PvdD.

Partijen die voor de eeuwwisseling zijn opgericht zijn gemiddeld genomen consistent. Daarnaast liggen deze partijen meer in het midden. Ten opzichte van de vorige Tweede Kamerverkiezingen zijn SP en CDA meer hun eigen verhaal gaan vertellen. Met minder goede wil zou men kunnen stellen dat deze partijen meer op een eiland zijn beland. 50PLUS en DENK laten een tegenovergestelde beweging zien, en staan t.o.v. de vorige verkiezingen juist beduidend meer in het talige centrum. Verder interessant is de toegenomen overeenkomst van de VVD met de PvdD en GroenLinks; de “GroenRechts-agenda van de VVD lijkt te zijn afgestoft. Ook is er sprake van een zekere mate van verwijdering tussen de coalitiepartijen. Tot slot hebben CU en D66 sinds 2017 hun middenposities verstevigd.

Door middel van deze uiteenzetting heb ik hopelijk drie zaken tentoon kunnen spreiden. Ten eerste heb ik laten zien dat politieke partijen zich in hun taal op verrassende wijze tot elkaar kunnen verhouden. Ten tweede heb ik laten zien hoe deze verhoudingen aan verandering onderhevig zijn. Ten derde heb ik de potentie laten zien van *Text Mining* als gereedschap voor onderzoek in het politieke domein.

## 6 Dankwoord

Zonder het meedenken en de steun van enkele zielen uit mijn omgeving had dit onderzoek niet tot stand komen. In het bijzonder erkentelijk ben ik de proeflezers van eerdere versies van dit artikel: Pieter van Casteren, Job Meijer, Koen Scholten, Quirine Smit en Dick Spruitenburg. Dankjulliewel voor jullie tijd en aandacht.

## Referenties

Andreas Hotho, Andreas Nurnberger, and Gerhard Paaß. A Brief Survey of Text Mining. pages 1–37, 2005.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. pages 1–8, 2020.

Hongyu Gong, Tarek Sakakini, Suma Bhat, and JinJun Xiong. Document Similarity for Texts of Varying Lengths via Hidden Topics. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2341–2351, Melbourne, Australia, 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1218. URL <http://aclweb.org/anthology/P18-1218>.

Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger. From word embeddings to document distances. In *Proceedings of the 32nd*

*International Conference on International Conference on Machine Learning - Volume 37*, ICML’15, page 957–966. JMLR.org, 2015.

## Verkiezingsprogramma's

50PLUS. De kracht van plus. Verkiezingsprogramma 2021-2025.

CDA. Nu doorpakken. Verkiezingsprogramma 2021-2025.

ChristenUnie (CU). Kiezen voor wat écht telt. Verkiezingsprogramma 2021-2025.

D66. Een nieuw begin. laat iedereen vrij, maar niemand vallen. Verkiezingsprogramma 2021-2025 d66.

DENK. Denk anders. Samen zijn wij nederland. DENK voor de toekomst. Verkiezingsprogramma denk 2021-2025.

FVD. Stem nederland terug. Verkiezingsprogramma 2021.

GroenLinks (GL). Tijd voor nieuw realisme. Verkiezingsprogramma groenlinks 2021.

PvdA. Ons plan voor een eerlijker en fatsoenlijker nederland. Verkiezingsprogramma pvda 2021-2025.

PvdD. Plan B: Idealisme is het nieuwe realisme. Verkiezingsprogramma Partij voor de Dieren Tweede Kamerverkiezingen 2021.

PVV. Het gaat om u. Verkiezingsprogramma 2021-2025.

SP. Stel een daad. Verkiezingsprogramma van de sp voor de Tweede Kamerverkiezingen van 17 maart 2021.

SGP. In vertrouwen. Verkiezingsprogramma 2021-2025.

VVD. Samen aan de slag. nieuwe keuzes voor een nieuwe tijd. Verkiezingsprogramma 2021-2025.