

# Duidelijke en eenvoudige taal. Hoe beoordeel je die?



De AVG wordt nu bijna twee jaar gehandhaafd en elke organisatie heeft inmiddels wel een privacystatement. Maar hoe zit het met de kwaliteit daarvan? In The Privacy Project van de New York Times (1) deden ze er al onderzoek naar. Hierdoor geïnspireerd vroegen we ons af hoe het in Nederland gesteld is met de privacyverklaringen. Is de kwaliteit hiervan tastbaar te maken? Waar moet je op letten? We doen met dit artikel een poging.

**O**ver de communicatie van bedrijven naar personen zegt de AVG (2): ‘... Voor natuurlijke personen dient het transparant te zijn dat hen betreffende persoonsgegevens worden verzameld, gebruikt, geraadpleegd of anderszins verwerkt en in hoeverre de persoonsgegevens worden verwerkt of zullen worden verwerkt. Overeenkomstig het transparantiebeginsel moeten informatie en communicatie in verband met de verwerking van die persoonsgegevens eenvoudig toegankelijk en begrijpelijk zijn, en moet duidelijke en eenvoudige taal worden gebruikt.’ Inderdaad, als het doel is om eenvoudig toegankelijk en begrijpelijk te zijn, is duidelijke en eenvoudige taal een goed middel. Meer staat er dan eigenlijk niet in de AVG. Deze tekst wordt nog een paar keer herhaald en daar blijft het bij. De privacyverklaring is niet de enige tekstsoort die de noodzaak heeft toegankelijk en begrijpelijk te moeten zijn. Dat moet gelden voor alle communicatie van de overheid. Over dit onderwerp moet dus wel meer informatie te vinden zijn.

### Duidelijke taal

Als we het hebben over eenvoudige taal, dan hebben we het in de wandelgangen over Jip-en-Janneke taal, maar dat is eigenlijk niet correct. Annie M.G. Schmidt was een expert in het beschrijven van uiteenlopende onderwerpen in zeer eenvoudige taal, geschikt voor zelfs de beginnende lezer. Het is moeilijk om op dit niveau volwassen onderwerpen zoals privacy goed genoeg uit te leggen. De

overheid heeft zelf de norm te communiceren met de burger op een hoger niveau, dat aangeduid wordt met het Europees Referentiekader-niveau B1 of het onderwijsniveau 2F. Dit komt overeen met het minimum taalniveau van iemand met een mbo-2/mbo-3 opleiding (3). Bevestiging dat de auteurs van de AVG er ook zo over denken komt in de volgende passage over communicatie specifiek gericht op kinderen:

‘... Aangezien kinderen specifieke bescherming verdienen, dient de informatie en communicatie, wanneer de verwerking specifiek tot een kind is gericht, in een zodanig duidelijke en eenvoudige taal te worden gesteld dat het kind deze makkelijk kan begrijpen.’

Het niveau van de tekst alleen is natuurlijk niet zaligmakend, daar is genoeg discussie over (4), maar het is desondanks een concreet fundament waar we mee kunnen werken. Dus werd het zaak om middelen te vinden die ons een indicatie geven van het ERK-niveau of het onderwijsniveau van een tekst. Dat bleek echter moeilijk. De beschrijvingen van de ERK-niveaus zijn kwalitatief en hebben het oordeel van een taalkundige nodig. Het CITO heeft een programma om de onderwijsniveaus te bepalen op basis van een Cito LeesIndex voor het Basisonderwijs-score, maar dat programma is niet publiekelijk beschikbaar. Uiteindelijk hebben we twee middelen gevonden die ons geholpen hebben de teksten op eenvoud en duidelijkheid te analyseren, zie het kader ‘Middelen’.

## Middelen

Het CITO-programma P-CLIB bepaalt het onderwijsniveau van een tekst door een leesbaarheidsscore te bepalen. De informatie over de bepaling van deze score is te summier en niet consistent. Wel wordt het duidelijk welke tekstenmerken gebruikt worden in die beoordeling (5):

- GWL = gemiddelde woordlengte in letters
- PFREQ = percentage hoogfrequente woorden
- TTR = type-token-ratio (aantal verschillende woorden gedeeld op het totaal aantal woorden)
- GZW = gemiddeld zinslengte per woord

Om een idee te geven: als we naar de volledige tekst van de Universele Verklaring van de Rechten van de Mens (UVRM) kijken (10984 letters, 1987 woorden, 95 zinnen), vinden we dat de gemiddelde woordlengte 5,52 letters is, en de gemiddelde zinslengte 20,9 woorden is. Het gebruikt 573 unieke woorden (28,8%). De meeste van deze kenmerken zijn objectief te meten voor een tekst via een aantal tools/services. Wij konden geen middel vinden dat specifiek gericht is op het Nederlands, dus hebben we ons gericht op tools voor de Engelse taal en kozen WebFX' Readability Test Tool (6). De uitzondering die moeilijker te meten valt is PFREQ, het percentage hoogfrequente woorden. Dit meetpunt zegt ook meer over de duidelijkheid van het taalgebruik, daarom schuiven we dit kenmerk door naar het andere aspect waar we naar moeten kijken.

## Tekstlengte

Eerst is er nog iets te zeggen over een ander kenmerk: de tekstlengte. Ook dit raakt de begrijpelijkheid, want een lange tekst doornemen kost meer tijd. Een (vlotte) lezer leest ongeveer 250 woorden per minuut. Een langere tekst vergt langere concentratie en vermoeit de lezer dus meer. Dit speelt voor de universele bepaling van eenvoud geen rol, maar in onze context wel. Net zoals een toespraak, hoort een privacyverklaring lang genoeg te zijn om alles te omvatten en kort genoeg om de aandacht vast te houden. De verklaring moet kort zijn, maar ook een bepaalde inhoud dekken en zal dus een bepaalde lengte moeten hebben. Hier is dus sprake van een optimum dat gevonden moet worden tussen deze twee aspecten. Eerder hebben we gezien dat de UVRM ongeveer 2000 woorden bevat. Onze verwachting is dat een privacyverklaring vergelijkbare eenvoudige taal moet hebben, met een vergelijkbare lengte en dus eenvoudig in tien minuten door te lezen is. En dit is nog maar het eerste aspect waarop we de gebruikte taal moeten beoordelen. Er moet ook nog duidelijke taal gebruikt worden. Hiervoor dien je naar het woordgebruik te kijken. Woorden zijn eenvoudig objectief te tellen en woordfrequenties ook. Dus ook hier gingen we op zoek naar een tool. We zijn uitgekomen op een initiatief van een Nederlandse wiskundedocent (7), vooral omdat de resultaten eenvoudig copy-paste-baar zijn. Op deze eenvoudige woordtellingen kunnen additionele analyses uitgevoerd worden om te bepalen hoe duidelijk de taal is, bijvoorbeeld door de waarde van het PFREQ-kenmerk te bepalen. Hiervoor is wel een lijst nodig van de meest gebruikte Nederlandse woorden. Deze is beschikbaar, maar er is geen formele standaardlijst. Als we een lijst kiezen en vastleggen is het percentage frequente woorden in een tekst objectief te bepalen. Een ander kenmerk dat in beschouwing genomen wordt in de WebFX Readability Test Tool is het aantal

lettergrepen per woord. Het gebruik van woorden met meer lettergrepen maakt de tekst minder begrijpelijk. Gevoelsmatig klinkt dit correct. Kijken we weer naar de UVRM, dan zien we dat (naar schatting) veel frequente woorden worden gebruikt, ongeveer 95%. Het percentage complexe woorden (drie lettergrepen of meer) is 25,2%. WebFX is geschikt voor de Engelse taal. We verwachten echter dat als we WebFX gebruiken om een lettergreetelling op een Nederlandse tekst uit te voeren we desondanks een betrouwbaar resultaat krijgen door de gelijkheid in de talen. Onze testen wijzen uit dat er zelfs een direct verband is tussen gemiddelde woordlengte en het percentage complexe woorden. Het is dus niet nodig de lettergreetelling apart mee te nemen. De duidelijkheid van een tekst bepalen blijkt dus moeilijker dan de eenvoud ervan bepalen, maar het is mogelijk dit redelijk objectief te doen, door te letten op het percentage unieke woorden per tekstlengte en de gemiddelde woordlengte. Daarnaast is de duidelijkheid ook gediend bij het gebruik van duidelijke woorden. Hierbij verwachten we geen vaagheden, minder algemene woorden en meer specifieke woorden. Voor het domein privacy en de context privacyverklaring denken wij dat de volgende klassen indicaties kunnen geven van groepen gerelateerde specifieke woorden:

- basistermen: persoonsgegevens, gegevens, informatie;
- online-activiteiten: profiel, account, social media, website, cookies;
- organisatieactiviteiten: verwerken, gebruiken, bewaren, opslaan;
- juridisch: privacy, wetgeving, recht(en), belang, toestemming.

### Hoe doen Nederlandse bedrijven het?

Met de middelen om privacyverklaringen te analyseren in de hand hebben we de gelegenheid om een aantal Nederlandse organisaties onder de loep te nemen. We doen niet zo'n breed onderzoek als de New York Times deed, we voeren slechts een steekproef uit. We kozen representanten uit verschillende sectoren. Zo kwamen we tot een lijst van zestien bedrijven:

- Supermarkten: Jumbo, Albert Heijn en Lidl;
- Onderwijsinstellingen: Universiteit van Amsterdam en Avans Hogeschool;
- Zorginstellingen: Erasmus MC en Parnassia Groep
- OV-bedrijven: Nederlandse Spoorwegen, Connexion en GVB;
- Banken: ING, Rabobank en ABN AMRO;
- Verzekeraars: Nationale Nederlanden, Achmea
- Winkelketen: IKEA

De privacyverklaring die we gebruikt hebben is de eerste volledige verklaring die we via een verwijzing op de homepage van de organisatie vonden. We hebben geen samenvattingen gekozen en geen deelverklaringen, als bijvoorbeeld een cookieverklaring apart aangeboden werd.

### Lengtekenmerken

Om de eenvoud van het taalgebruik te beoordelen kijken we naar de lengtekenmerken: woordlengte, zinslengte en tekstlengte. Deze elementen zijn opgenomen in tabel 1.

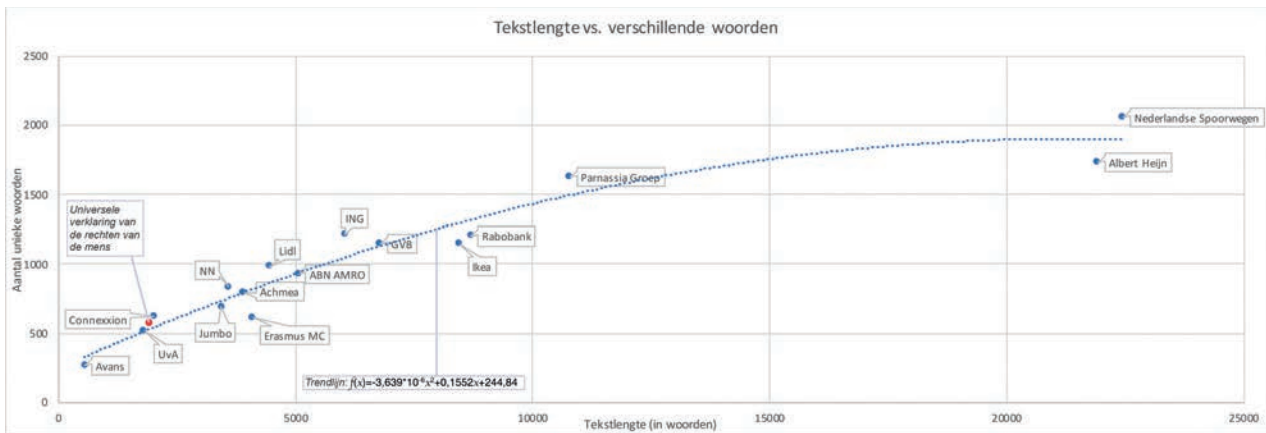
### Woordlengte

Als we kijken naar de woordlengte van de bekeken privacyverklaringen, dan is er niet zoveel verschil te zien. Dit varieert van 4,9 tot 6,2. Binnen deze bandbreedte valt wel op dat onderwijsinstellingen en gezondheidszorg langere woorden gebruiken. De teksten hebben een gemiddelde woordlengte van 5,9 letters. Twee voorbeelden uit de privacyverklaringen van Avans en Parnassia om een gevoel te geven hoe dat leest: 'Avans Hogeschool draagt zorg voor een adequaat beveiligingsniveau. Er is voorzien in passende technische en organisatorische maatregelen om persoonsgegevens te beveiligen tegen verlies of tegen enige vorm van onrechtmatige verwerking.' (6,8 letters per woord).

'De gegevensverwerking met het oog op archivering in het algemeen belang, wetenschappelijk of historisch onderzoek of statistische doeleinden is onderworpen aan passende waarborgen in overeenstemming met de AVG voor de rechten en vrijheden van de betrokkene.' (6,1 letters per woord).

Bedrijf	Branch	Eenvoudige taal				Duidelijke taal			
		Leestijd (min.)	Tekstlengte	Woordlengte	Woorden per zin	Verschillende woorden	Afwijking van trendlijn	Percentage complex	
Jumbo	Supermarkt	14	3427	5,5	17,0	692	-6%	25%	
Albert Heijn	Supermarkt	88	21915	5,2	12,4	1732	-9%	21%	
Lidl	Supermarkt	18	4436	5,5	14,4	985	14%	23%	
Universiteit van Amsterdam	Onderwijs	7	1800	5,9	17,1	517	1%	29%	
Avans Hogeschool	Onderwijs	2	545	6,2	15,9	262	-20%	31%	
Erasmus MC	Gezondheidszorg	16	4097	5,6	18,9	615	-25%	28%	
Parnassia Groep	Gezondheidszorg	43	10772	6,0	18,6	1627	9%	29%	
Nederlandse Spoorwegen	Openbaar vervoer	90	22452	5,5	17,3	2062	9%	26%	
Connexion	Openbaar vervoer	8	2010	5,9	15,4	622	15%	29%	
GVB	Openbaar vervoer	27	6776	5,8	20,2	1152	2%	26%	
ING	Financiële dienst	24	6050	5,4	14,5	1213	15%	26%	
Rabobank	Financiële dienst	35	8695	5,4	12,4	1201	-9%	25%	
ABN AMRO	Financiële dienst	20	5055	5,4	13,4	931	-1%	23%	
Nationale Nederlanden	Verzekeraar	14	3583	5,7	13,4	834	11%	26%	
Achmea	Verzekeraar	16	3890	5,7	16,0	791	0%	27%	
Ikea	Winkelketen	34	8454	4,9	15,8	1148	-11%	19%	
<i>Universele verklaring van de rechten van de mens</i>		<i>Voorbeeld</i>	8	1987	5,5	20,9	573	6%	25%
		gemiddelde	28	6820	5,6	16,1	997	NVT	26%
		minimum	2	545	4,9	12,4	262	-25%	19%
		maximum	90	22452	6,2	20,9	2062	15%	31%
		variatie	190,5%	NVT	22,8%	50,8%	NVT	NVT	48,8%

Tabel 1 - Overzicht Eenvoudskennmerken.



Figuur 1 - Scatter-diagram Lengte vs. Unieke woorden.

Supermarkten en banken gebruiken kortere woorden. IKEA spant de kroon met een gemiddelde woordlengte van slechts 4,9 letters. Lees ter vergelijking een passage uit hun verklaring: 'We gebruiken aankoopgegevens om beter inzicht te krijgen in onze klanten en de manier waarop ze winkelen. Dit doen we om het aanbod in onze winkel te verbeteren. De gegevens voor deze analyses bewaren we twee jaar.' (4,7 letters per woord).

De zinslengte geeft een afwisselender beeld, deze waarde varieert van 12 tot 20 woorden. Dit is opvallend, omdat in dit geval de privacyverklaringen allemaal onder de gemiddelde zinslengte zitten van de verklaring van de rechten van de mens (20,9 woorden per zin). Kennelijk is het gebruik van korte zinnen een eenvoudig toepasbare maatregel om taal te vereenvoudigen.

### Tekstlengte

Bij tekstlengte zien we grote verschillen. De verklaring van de rechten van de mens mag in tien minuten te lezen zijn, weinig privacyverklaringen halen dat: alleen die van de UvA, Avans en Connexion. Neem voor de gemiddelde privacyverklaring maar een half uur en voor de langste anderhalf uur (Albert Heijn en de Nederlandse Spoorwegen). Misschien zijn die lange versies niet bedoeld om van voor tot achter door te lezen. Zo heeft Albert Heijn een interactieve versie op haar website staan, waar je kunt kiezen over welk onderwerp je meer wil lezen. Anderen kiezen bijvoorbeeld om een samenvatting naast de verklaring aan te bieden, zoals de Parnassia Groep en het GVB. Voor de meeste verklaringen duurt het lezen ervan flink langer dan het lezen van de verklaring van de rechten van de mens, tussen een kwartier en een ruim half uur. Dat is te doen, maar je moet er wel voor gaan zitten. We vermoedden al dat er een afweging gemaakt moest worden tussen woordlengte en tekstlengte. De korte verklaringen hebben relatief een langere woordlengte. Over het geheel heeft IKEA een mooie prestatie geleverd met een hele korte woordlengte bij een tekstlengte net boven gemiddeld (+19%).

### Duidelijke taal

Om de duidelijkheid van de taal in de privacyverklaringen te bepalen kijken we naar het gebruik van verschillende woorden en specifiek woordgebruik (zie kader 'Middelen'). In een verklaring kun je verwachten dat het aantal unieke woorden hoger is naarmate de tekst langer is. Tegelijkertijd zal de groeitrend hiervan afnemen bij langere teksten, zie figuur 1. Deze figuur toont hoe de bekeken verklaringen voldoen aan deze verwachting door teksten te plotten op basis van hun lengte en aantal unieke woorden.

Het was mogelijk om over de punten in de figuur statistisch een passende curve te vinden waar de bekeken verklaringen minimaal van afwijken. Op dit punt zijn de teksten kennelijk goed vergelijkbaar. Teksten boven de curve gebruiken relatief meer verschillende woorden (bijvoorbeeld Lidl, Connexion, ING), teksten onder de curve gebruiken relatief minder verschillende woorden (bijvoorbeeld Avans Hogeschool en Erasmus MC). Mogelijk is de verklaring hiervoor dat boven de curve de meer gevarieerde teksten staan en onder de curve meer gestructureerde teksten.

Als laatste kijken we naar specifiek woordgebruik en de consistentie daarvan. We hebben vier klassen van specifieke woorden gedefinieerd (zie kader 'Middelen') en bekijken hoe de verklaringen woorden uit die klassen gebruiken. Dit geeft ons de duidelijkheid van de taal in de context van een privacyverklaring en daarmee dus inzicht in de helderheid van die verklaring.

In vrijwel alle privacyverklaringen vallen de basistermen 'gegevens' en 'persoonsgegevens' onder de meest voorkomende woorden. Persoonsgegevens slaat echter op geïnterpreteerde of interpreteerbare gegevens, dus informatie. Lees wat persoonsgegevens nu precies zijn, volgens de AVG (2): '... alle informatie over een geïdentificeerde of identificeerbare natuurlijke persoon ...' De term 'gegevens' lijkt dus juist niet duidelijk genoeg. Dat het toch overwegend gebruikt wordt, komt wellicht doordat deze basisterm als definitie is overgenomen uit de wet. Opvallend genoeg

Bedrijf	Gegevens en informatie			Social media en online				Organisatie	Juridisch					
	Gegevens	Informatie	Persoons-gegevens	(gebruikers)a Profiel(en)	Social(e) + ccount	media	website		cookies	Verwerken Verwerking	privacy	recht(en)	belang	toestem- ming
Jumbo	6	88	22	2	25	x	x	8	9	24	15	10	8	5
Albert Heijn	304	88	27	169	8	11	79	13	114	37	6	7	22	1
Lidl	47	13	37	x	12	11	37	54	12	17	12	6	15	5
Universiteit van Amsterdam	7	7	70	x	1	1	4	1	21	7	1	4	7	5
Avans Hogeschool	6	1	9	x	x	x	1	x	4	5	2	2	x	2
Erasmus MC	65	7	95	x	x	x	1	1	61	10	11	10	9	13
Parnassia Groep	71	44	150	x	x	3	2	x	87	10	29	17	27	38
Nederlandse Spoorwegen	377	83	190	2	14	6	28	32	110	68	16	39	51	13
Connexion	30	7	25	x	x	6	8	8	9	17	3	3	5	2
GVB	48	20	122	x	7	3	5	3	27	13	10	3	7	8
ING	76	26	63	7	x	2	5	9	20	54	29	3	10	10
Rabobank	196	19	78	4	x	4	12	3	53	44	21	27	20	15
ABN AMRO	19	5	124	6	x	3	12	3	8	17	14	11	11	24
Nationale Nederlanden	66	7	54	1	2	x	5	3	19	20	10	5	6	7
Achmea	90	5	11	x	x	2	5	2	10	20	12	5	4	9
Ikea	175	19	9	40	33	17	20	42	12	54	8	7	15	4

Tabel 2 - Overzicht Wordfrequenties.

lijken de communicatiemedewerkers bij Jumbo het met ons eens te zijn. Zij prefereren als enige duidelijk het woord 'informatie' (88 keer gebruikt) boven 'gegevens' (6 keer gebruikt).

Veel tekst in verklaringen heeft betrekking op online-activiteiten. De supermarkten Albert Heijn en Lidl en winkelketen IKEA verwijzen vaak naar woorden in deze klasse. Komt dit doordat hun aandacht voor commerciële online-activiteiten de laatste jaren flink zijn toegenomen? Dit is waarschijnlijk in mindere mate ook het geval bij de Nederlandse Spoorwegen en Connexion. De banken doen veel online-activiteiten, maar daar is het gebruik meer essentieel dan commercieel. Wellicht heeft dit daarom voor hen geen specifieke aandacht nodig in de privacyverklaring.

Onder de organisatieactiviteiten valt vooral op dat een privacyverklaring geschreven is met vormen van het woord 'verwerken' (Albert Heijn, Parnassia Groep, Nederlandse Spoorwegen, Rabobank) of juist alternatieven daarvan ('gebruiken', 'bewaren', 'opslaan'). Zien we hier nog een schim van de jurist in de tekst van de verklaring?

Woorden in de juridische klasse scoren over het geheel genomen laag. Dit geeft aan dat men kennelijk de moeite heeft genomen om de formele, juridische termen niet te vaak te noemen. Albert Heijn spant hier de kroon, buiten het woord 'privacy' worden slechts 36 juridische termen gebruikt in hun lange verklaring.

Om de duidelijkheid van een privacyverklaring te beoordelen is het kijken naar het woordgebruik het meest bepalend. Een duidelijke keuze maken en dan consistent zijn met het woordgebruik is waarschijnlijk de beste strategie. Dan gaat ook het percentage unieke woorden ten opzichte van de tekstlengte omlaag, wat verder bijdraagt aan de duidelijkheid.

## Bevindingen

Harde conclusies kunnen we niet trekken op basis van wat we gezien hebben met deze kleine steekproef onder privacyverklaringen. Wel kunnen we zeggen dat het aantoont dat het best goed gesteld is met de leesbaarheid van deze teksten. Zonder de eenvoud taalkundig te hebben laten beoordelen lijken de tekstkenmerken overeen te komen met ERK-niveau B1, soms wat lager (IKEA) en soms wat hoger (Universiteit van Amsterdam en Avans Hogeschool). Voor wat betreft de duidelijkheid is het taalgebruik consistent, soms wat formeel, zoals het gebruik van de wetsterm 'verwerken'. Een grote uitzondering vonden we in de variëteit in lengte. Meer vergelijkingen zijn nodig om te zien of dit een afwijking is in de steekproef, of dat deze verscheidenheid doorzet. Meer vergelijkingen geven ook meer inzicht in de trends in het woordgebruik. Met name de sectoren die nog niet opgenomen zijn in onze steekproef (zoals overheden, industrie, MKB) kunnen wellicht geheel nieuwe sectorale inzichten verschaffen.

## Referenties

- (1) New York Times - We Read 150 Privacy Policies. They Were an Incomprehensible Disaster: <https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html>
- (2) AVG: [https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/verordening\\_2016\\_-\\_679\\_definitief.pdf](https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/verordening_2016_-_679_definitief.pdf).
- (3) Staatsexamens Nf2: <https://www.staatsexamensnt2.nl/veelgevraagd/met-welk-cefr-erk-niveau-komen-de>
- (4) Volkskrant - Pleidooi voor duidelijke taal niet altijd even helder: <https://www.volkskrant.nl/gsb1a4332e>.
- (5) Begrijpelijke taal - Begripsvoorspelling: [http://www.kennisbank-begrijpelijketaal.nl/begripsvoorspelling/ned\\_formules](http://www.kennisbank-begrijpelijketaal.nl/begripsvoorspelling/ned_formules).
- (6) Readability Test Tool: <https://www.webfx.com/tools/read-able/>.
- (7) Woord en letterteller: <http://home.wxs.nl/~hklein/zipf/zipf.htm>