

Leesindex A, leesindex B en leesindex AB voor de bepaling van de leesbaarheid van Nederlands proza

R. H. M. BROUWER

Samenvatting

Reeds in 1963 is in dit tijdschrift de bepaling van de leesbaarheid van Nederlands met behulp van de gemiddelde woord- en zinslengte (leesindex A) beschreven. In 1972 volgde het onderzoek vanuit de factor woordkeus (leesindex B).

In deze aflevering worden beide onderzoeken met elkaar in verband gebracht. Dit levert de leesindex AB-formule op. Op grond van een correlatie ($r = 0.854$) tussen leesindex A en leesindex AB lijkt het verantwoord de leesbaarheid van Nederlandse teksten te bepalen met behulp van de leesindex A. Het is dan toch mogelijk ook iets over de factor woordkeus te zeggen. Voor de interpretatie van de leesindex A-cijfers dient men wel de oorspronkelijk bij leesindex B behorende normtabel te hanteren.

Uit de onderzoeksgegevens van leesindex A is ook nog een correlatie ($r = -0.99$) afgeleid tussen de gemiddelde woordlengte en het percentage eenlettergrepige en tweelettergrepige woorden samen. Dit levert een dubbele besparing op, t.w. minder telwerk en een eenvoudiger leesindex A-formule.

Leesindex A

Leesindex A bepaalt de leesbaarheid van een tekst gezien vanuit de factoren gemiddelde woord- en zinslengte. In het oktobernummer van 1963 van dit tijdschrift is het onderzoek beschreven (Brouwer, 1963).

De problemstelling luidde: Is in het Nederlands proza de gemiddelde zinslengte – uitgedrukt in woorden – alsmede de gemiddelde

woordlengte – uitgedrukt in lettergrepen – van invloed op de leesbaarheid; en zo ja, is het mogelijk een gradatie in moeilijkheid te bepalen?

Als criterium is de volgende indeling van de Nederlandse lectuur genomen: kinderlectuur 6–7–8 jaar, idem 9–10–11 en 12 jaar, idem 13–14–15 jaar, specifieke jeugdlectuur, detectives, cowboys, damesromans, reisverhalen, historische romans, streekromans, populair wetenschappelijke boeken en studieboeken. Van elke categorie zijn twee representatieve boeken doorgemeten (alleen bij de eerste groep is één – volgens vele deskundigen zeer makkelijk boek – toegevoegd), in totaal dus 25 boeken. Per boek werd – m.b.v. een streekproef van $7 \times \pm 300$ woorden – de gemiddelde woord- en zinslengte berekend.

Gebleken is, dat inderdaad – in het algemeen – de gemiddelde zins- en woordlengte stijgen naarmate een boek moeilijker leesbaar wordt geacht. De gemiddelde zinslengte van de onderzochte boeken varieerde van 6 t/m 27, de gemiddelde woordlengte van 1,33 tot en met 1,96.

Daar uit het doormeten van incidentele teksten was gebleken, dat er ook hogere c.q. lagere waarden voor de gemiddelde woord- en zinslengte voorkomen en om een handzame formule te creëren, zijn als gemiddelde uiterste waarden genomen 5 en 30 voor de zinslengte en 1,27 en 2,02 voor de woordlengte.

Gesteld werd dat:

- bij een zo laag mogelijk gemiddeld aantal woorden per zin en een zo laag mogelijk gemiddeld aantal lettergrepen per woord de leesmoeilijkheid het kleinst is ($100 = a + 1,27b + 5c$);
- bij een zo hoog mogelijk gemiddeld aantal

woorden per zin en een zo hoog mogelijk gemiddeld aantal lettergrepen per woord de leesbaarheid het grootst is ($0 = a + 2,02b + 30c$);

- c. voorlopig voor de tussenliggende waarden een lineaire verdeling bestaat met dien verstande dat evenveel invloed is toegekend aan de totale variatie in de gemiddelde zinslengte als aan die in de gemiddelde woordlengte ($50 = a + 1,27b + 30c$ en $50 = a + 2,02b + 5c$).

Op grond van de gemiddelde woord- en zinslengten en deze stellingen komt men tot de volgende formule:

leesindex A = $195 - 2/3$ woordlengte - 2 zinslengte (waarbij de woordlengte wordt vermenigvuldigd met 100).

Met behulp van de gevonden nauwkeurigheden kon door samenvoeging van verschillende categorieën en door afronding een indeling naar moeilijkheid worden vastgesteld. Voor interpretatie van het leesindex A-cijfer dient men tabel 2 (blz. 5) te raadplegen.

Leesindex B

Analoog aan het onderzoek naar de gemiddelde zins- en woordlengte (leesindex A) is een leesbaarheidsformule ontwikkeld op basis van de woordkeus (leesindex B). Dit onderzoek is beschreven in het meinummer 1972 van Pedagogische Studiën (Brouwer, 1972).

Hierbij is van dezelfde indeling Nederlandse lektuur uitgegaan met dien verstande dat de categorie studieboeken en het toegevoegde 'zeer makkelijke boek' kwamen te vervallen.

De 22 boeken zijn doorgemeten m.b.v. een steekproef van $9 \times \pm 300$ woorden. Van alle woorden uit de verschillende steekproeven is nagegaan of zij voorkomen in De la Courts 'Lijst van de in het Nederlandsch meest voorkomende en meest bruikbare woorden'. De niet in deze lijst voorkomende woorden zijn 'outsiders'. Van deze outsiders is tevens de frequentie nagegaan. De meermalen voorkomende outsiders zijn om-

gezet in herleide outsiders. De herleide outsiders en de eenmaal voorkomende outsiders vormen samen de 'gewogen' outsiders. Het minimum percentage 'gewogen' outsiders bedraagt 2,105 en het maximum 10,688. Evenals bij leesindex A is een marge aangehouden. Vandaar dat het laagste en het hoogste percentage gewogen outsiders bepaald zijn op respectievelijk 1 (i.p.v. 2,105) en 13,5 i.p.v. 10,688). Door evenals bij leesindex A een lineair verband te veronderstellen, komt men tot de formule:

$$y = ax + b,$$

waarbij

$$100 = a \times 1 + b$$

en

$$0 = a \times 13,5 + b.$$

Dit leidt tot de formule:

$$\text{leesindex B} = 108 - 8x$$

(waarin x = percentage gewogen outsiders).

Met behulp van de gevonden nauwkeurigheden kon ook hier door samenvoeging van verschillende categorieën en door afronding een indeling naar moeilijkheid worden vastgesteld. De waarde van de leesindex B-cijfers is af te lezen in tabel 2 (blz. 5).

Leesindex AB

Een vergelijking van het onderzoek naar de leesbaarheid van Nederlands proza met behulp van de leesindex A (gemiddelde woord- en zinslengte) en leesindex B (woordkeus) laat zien, dat beide indexcijfers niet per se per boek van gelijke waarde hoeven te zijn. Tevens blijkt, dat de benutte indeling Nederlandse literatuur niet de opklimming in moeilijkheid volledig weergeeft. De streekromans bijvoorbeeld moeten gerekend worden tot makkelijk te lezen literatuur. Verder zijn er boekengenres, die in aanmerking komen voor plaatsing in meer dan één categorie. Bijvoorbeeld 'Kampvuren langs de evenaar' van Dr. Paul Juliën kan men zowel plaatsen bij reisverhalen, maar ook bij het genre populair-weten-

Tabel I. Rangvolgorde van de boeken 1 t/m 22 op grond van leesindex AB.

Schrijver	Leesindex			Categorie*
	AB	A	B	
2 Frans Franssen	83	81	85	I 100-76 gemakkelijk
3 A. D. Hildebrand	82	73	91	
4 To Hölscher	81	80	80	
1 Dick Laan	77	69	85	
9 Agatha Christie	76	77	74	II 80-56 standaard
12 Wm. Colt Macdonald	72	80	61	
8 Sanne v. Havelte	71	75	66	
5 J. Nowee	70	69	71	
14 Warwick Deeping	69	71	66	
11 Max Brand	68	73	64	
10 Dorothy L. Sayers	66	74	57	
6 Hella Jansonius	65	72	56	
7 Fritz Steuben	65	69	59	
19 Herman de Man	62	73	51	
13 Margarete Elzer	61	65	56	
20 Marie Koenen	58	62	52	
16 Mary Pos	44	37	51	III 60-36 tamelijk moeilijk
17 Hella S. Haasse	42	55	28	
15 Dr. Paul Juliën	36	36	36	
18 Jo v. Ammers-Küller	34	34	33	IV 40-16 moeilijk
21 Anton Zischka	32	25	38	
22 Mr. H. C. M. Edelman	31	40	22	
23 J. M. J. Korpershoek		20		V 20-0 zeer moeilijk
24 Prof. Dr. F. J. J. Buytendijk		16		

*Een zekere overlapping van de categorieën is onvermijdelijk.

schappelijk.

Het is mogelijk een formule voor de leesbaarheid samen te stellen, waarin leesindex A en leesindex B even zwaar vertegenwoordigd zijn¹⁾.

Deze formule luidt:

leesindex AB = 151,5 - 1/3 woordlengte - zinslengte - 4 gewogen outsiders.

Op grond van de leesindex AB is een nieuwe volgorde van toeneming in moeilijkheid ontstaan, die in tabel I is weergegeven.

¹⁾De statistische verwerking is verzorgd door dhr. W. Knops, afd. Personeelsresearch Staatsmijnen.

Met behulp van de rangcorrelatie formule van Spearman is de correlatie tussen leesindex AB en leesindex A, en tussen leesindex AB en leesindex B bepaald. De eerste blijkt $r = 0.854$ en de tweede $r = 0.965$ te bedragen.

De tweede correlatie is zeer hoog te noemen, maar de berekening van leesindex B is een zo tijdrovende en ingewikkelde aangelegenheid dat deze formule niet op grote schaal zal worden toegepast.

Het lijkt dan ook verantwoord ($r = 0.854$) om in eerste instantie de leesbaarheid van een tekst te bepalen met behulp van de leesindex A-formule

(gemiddelde woord- en zinslengte). Mede omdat via deze ingang er toch wat over de woordkeus is te zeggen. Immers de hoge correlatie tussen leesindex AB en leesindex A, alsmede tussen leesindex AB en leesindex B gevoegd bij de hierna vermelde zeer hoge correlatie ($r = -0.99$) tussen het percentage één- plus twee-lettergrepige woorden en de gemiddelde woordlengte wettigen de conclusie, dat voor het Nederlands de factor woordkeus voornamelijk wordt bepaald door de drie- en meer lettergrepige woorden. Deze conclusie komt overeen met de gegevens van Zipf (1935) die zegt 'there exists a tendency in Peiping Chinese, Plautine Latin, and American and British English for the length of a word to be negatively correlated with its frequency of usage'.

Gezien vanuit de factor minder gebruikte woorden dient men dus aandacht te besteden aan de drie- en meerlettergrepige woorden. In verband met de geringere woordenschat van kinderen is het gewenst bij teksten voor kinderen ook aandacht te besteden aan de één- en tweelettergrepige woorden van buitenlandse origine (zie ook bevindingen voor het Engelse taalgebied in artikel literatuurlijst nr. 2), alsmede aan de minder bekende één- en tweelettergrepige woorden.

Op grond van de leesindex AB-gegevens is het gewenst om voor de interpretatie van de leesindex A-, B- en AB-cijfers de oorspronkelijk bij leesindex B behorende tabel te gebruiken.

Vereenvoudiging leesindex A-formule

Wil een leesbaarheidsformule veelvuldige toepassing vinden dan moet zij zo eenvoudig mogelijk zijn. In navolging van het werk van James, Jenkins en Paterson (1951) is naar een correlatie gezocht tussen de gemiddelde woordlengte en het percentage éénlettergrepige woorden.

In tabel 3 is een overzicht gegeven van de percentages woorden met één, twee, drie etc. lettergrepen bij de 25 schrijvers(sters). Deze gegevens tonen aan, dat er een zeer hoge correlatie ($r = -0.967$) bestaat tussen het percentage éénlettergrepige woorden en de gemiddelde woordlengte. Maar uit nadere bestudering bleek, dat er voor het Nederlands een nog hogere correlatie ($r = -0.99$) geldt voor het percentage één- plus tweelettergrepige woorden samen en de gemiddelde woordlengte. Deze correlaties bleken statistisch significant verschillend te zijn. De formule hiervoor luidt:

$$y = -3x + 411$$

(waarin y = gemiddelde woordlengte in lettergrepen $\times 100$ en x = percentage één- plus tweelettergrepige woorden samen).

Door dit verband wordt de bepaling van de gemiddelde woordlengte sterk vereenvoudigd. Immers in het onderzoek naar leesindex A verte-

Tabel 2. Leesmoeilijkheidsnormen leesindex A, B en AB.

Categorie	Leesmoeilijkheid		Lectuur-genre
	Index A	Omschrijving	
I	100-76	gemakkelijk	kinderlectuur t/m 12 jaar
II	80-56	standaard	kinderlectuur vanaf 13 jaar specifieke jeugdlectuur-detectives-cowboys- damesromans
III	60-36	tamelijk moeilijk	streekromans-reisverhalen
IV	40-16	moeilijk	historische romans-populair wetenschappelijk
V	20-0	zeer moeilijk	wetenschappelijke lectuur

Tabel 3. Overzicht van de percentages woorden met één, twee, drie enz. lettergrepen bij de schrijvers(-sters) ter bepaling van de correlatie tussen het percentage één- plus tweelettergrepige woorden samen en de gemiddelde woordlengte.

schrijver-ster	aantal lettergrepen per woord							%	gemiddelde woordlengte
	1	2	3	4	5	6	7		
1	75	18	7	1	—	—	—	93	133
2	69	21	10	—	—	—	—	90	143
3	71	21	6	2	—	—	—	92	140
4	68	24	6	1	—	—	—	92	140
5	69	23	7	2	—	—	—	92	142
6	63	24	10	2	—	—	—	87	153
7	65	23	10	2	—	—	—	88	151
8	65	25	8	2	—	—	—	90	149
9	70	21	8	1	—	—	—	91	141
10	64	25	8	2	1	—	—	89	150
11	65	23	9	2	1	—	—	88	151
12	68	22	7	3	1	—	—	90	147
13	69	21	9	1	—	—	—	90	142
14	62	24	8	5	1	—	—	86	158
15	66	23	8	2	1	—	—	89	150
16	57	23	13	5	2	—	—	80	172
17	60	24	11	4	1	—	—	84	165
18	61	23	12	4	1	—	—	85	161
19	62	23	11	4	1	—	—	85	160
20	69	21	7	2	—	—	—	90	143
21	63	24	9	3	1	—	—	87	154
22	56	22	11	7	2	1	1	78	183
23	57	20	13	6	2	1	—	77	181
24	52	22	13	8	3	2	—	74	196
25	55	19	13	9	3	1	—	74	190

genwoordigen de drie- en meerlettergrepige woorden slechts 14% van alle woorden (de percentages varieerden van 7 tot 26). Door het aantal drie- en meerlettergrepige woorden te tellen en dit aantal af te trekken van het totaal aantal woorden, weet men het aantal één- plus tweelettergrepige woorden. Dit aantal kan men daarna omzetten in een percentage van het totaal aantal woorden.

Tevens kan men het gegeven dat de gemiddelde

woordlengte (y) gelijk is aan $411 - 3x$ (waarbij x = percentage één- plus tweelettergrepige woorden samen) invullen in de leesindex A-formule.

Dit geeft dan het volgende te zien:

$$\text{leesindex A} = 195 - 2/3 \text{ wl} - 2\text{zl of}$$

$$\text{leesindex A} = 195 - 2/3(411 - 3x) - 2\text{zl of}$$

$$\text{leesindex A} = 195 - 274 + 2x - 2\text{zl of}$$

$$\text{leesindex A} = 2x - 79 - 2\text{zl.}$$

Voor de verschillende waarden van x is nu de gemiddelde woordlengte (y) en de waarde van de term $2x - 79$ te berekenen en in een tabel weer te geven. Voor de bepaling van de leesindex A hoeft men dan slechts 2zl van deze term af te trekken.

Tabel 4. Voor aflezing van gemiddelde woordlengte (y) en de leesindex A als percentage 1+2 lettergrepige woorden (x) en gemiddelde zinslengte (zl) bekend zijn.

x	y	leesindex A	x	y	leesindex A
100	111	121-2 zl	80	171	81-2 zl
99	114	119-2 zl	79	174	79-2 zl
98	117	117-2 zl	78	177	77-2 zl
97	120	115-2 zl	77	180	75-2 zl
96	123	113-2 zl	76	183	73-2 zl
95	126	111-2 zl	75	186	71-2 zl
94	129	109-2 zl	74	189	69-2 zl
93	132	107-2 zl	73	192	67-2 zl
92	135	105-2 zl	72	195	65-2 zl
91	138	103-2 zl	71	198	63-2 zl
90	141	101-2 zl	70	201	61-2 zl
89	144	99-2 zl	69	204	59-2 zl
88	147	97-2 zl	68	207	57-2 zl
87	150	95-2 zl	67	210	55-2 zl
86	153	93-2 zl	66	213	53-2 zl
85	156	91-2 zl	65	216	51-2 zl
84	159	89-2 zl	64	219	49-2 zl
83	162	87-2 zl	63	222	47-2 zl
82	165	85-2 zl	62	225	45-2 zl
81	168	83-2 zl	61	228	43-2 zl

Voor interpretatie van de leesindex A-cijfers zie tabel 2 (blz. 5)

Literatuur

Brouwer, R. H. M., Onderzoek naar de leesbaarheid van Nederlands proza; *Pedagogische Studiën*, jrg. 40 nr. 10 (okt. '63).

Brouwer, R. H. M., Verstaanbaarheid en meetbaarheid van de taal; *Dux*, jrg. 33 (juli '66).

Brouwer, R. H. M., Onderzoek naar de leesbaarheid

van het Nederlands gezien vanuit de factor woordkeus; *Pedagogische Studiën*, jrg. 49, nr. 5 (mei '72).

James, F. N., James H. Jenkins and Donald G. Paterson, Simplification of Flesch Reading Ease Formula; *Journal of Applied Psychology*, vol. 35, 1951, p. 333-337.

Zipf, G. K., *The psycho-biology of language*; Boston, Houghton, Mifflin, 1935.