

# Onderzoek naar de leesbaarheid van het Nederlands gezien vanuit de factor woordkeus

R. H. M. BROUWER\*

## Samenvatting

Ten vervolge op een in 1963 gepubliceerd artikel over leesbaarheid bepaald door de factor gemiddelde woord- en zinslengte, volgt nu een benadering vanuit de factor woordkeus. Ook bij dit onderzoek is uitgegaan van een indeling van de Nederlandse lectuur. Per onderscheiden lectuurgenre zijn van twee hiervoor karakteristieke schrijvers één boek bekeken. De boeken zijn doorgemeten m.b.v. een a-selecte steekproef per boek. Van deze steekproeven is nagegaan welke woorden vermeld zijn in de 'woordfrequentielijst' van De la Court e.a. De niet in de lijst voorkomende woorden werden omgezet in gewogen 'outsiders' op grond van het gegeven dat een outsider die verschillende malen in een tekst voorkomt niet steeds dezelfde moeilijkheidsgraad voor de lezer behoudt. Om de nauwkeurigheid van de indexcijfers te kunnen bepalen zijn 95%-betrouwbaarheidsgrenzen berekend. Met behulp van deze nauwkeurigheden kon – door samenvoeging van verschillende categorieën alsmede toepassing van afronding – een indeling naar moeilijkheid worden vastgesteld.

## 1. Inleiding

Dit artikel beschrijft een onderzoek naar de moeilijkheid van Nederlands proza, benaderd vanuit de factor woordkeus (leesindex B). In 1963 is in het oktobernummer van Pedagogische Studiën een soortgelijk onderzoek beschreven waarbij de

\* Bij dit onderzoek hebben wij zeer waardevolle adviezen ontvangen van de heer F. Timmers, Afdeling Statistiek N.V. Nederlandse Staatsmijnen

leesbaarheid bepaald wordt door de factor gemiddelde woord- en zinslengte (leesindex A).

## 2. Methode

Van een tweeëntwintigtal schrijvers – die karakteristiek zijn te noemen voor te onderscheiden lectuur-genres is één boek met behulp van steekproeven doorgemeten. Van alle woorden uit de verschillende steekproeven is nagegaan of deze behoren tot de woordfrequentielijst De la Court. De niet daarin voorkomende woorden worden 'outsiders' genoemd (met uitzondering van de eigennamen). Daar een 'outsider' als hij verschillende malen in een tekst voorkomt niet steeds dezelfde moeilijkheidsgraad voor de lezer behoudt, is een wegingsfactor ingevoerd. Het percentage 'gewogen outsiders' bepaalt de leesbaarheid van de tekst gezien vanuit de factor woordkeus (leesindex B). Om de nauwkeurigheid van de indexcijfers te kunnen aangeven zijn de 95%-betrouwbaarheidsgrenzen berekend. Met behulp van deze nauwkeurigheden kon, door samenvoeging van verschillende categorieën alsmede door toepassing van afronding, een indeling naar moeilijkheid worden vastgesteld.

## 3. Het woordfrequentie-onderzoek van De la Court e.a.<sup>2</sup>

Bij besluit van de Directeur van Onderwijs en Eeredienst van 14 september 1934 nr. 29178/B werd een commissie ingesteld, die tot taak had na te gaan of en zo ja op welke wijze de aan 'Basic-English' ten grondslag liggende beginse-

len aan de taalmethodiek op meisjesscholen en eventueel op standaardscholen dienstbaar zouden kunnen worden gemaakt, ten aanzien van het onderricht in de Nederlandse taal.

Het doel van het onderzoek van De la Court was een algemene woordenlijst samen te stellen; een woordenlijst die dus een beeld geeft van de gemiddelde taal en die zowel bruikbaar zou zijn voor het lezen als voor het spreken van de Nederlandse taal\*. De woordenlijst van De la Court is met medewerking van personeel en leerlingen van een kweekschool in Nederlands Indië in streng geleid groepswerk tot stand gekomen. Bij dit woordfrequentie-onderzoek zijn 320 teksten met een totaal aantal van 1.000.000 woorden onderzocht. Het 'algemeen' Nederlands werd gekozen uit teksten met een verschillende graad van moeilijkheid, beginnende bij de kinderlectuur en opklimmende tot de gewone romanlectuur. Met de behoeften van de spreektaal werd rekening gehouden door in de onderzochte teksten veel dialoog op te nemen. Oorspronkelijke lectuur en goede vertalingen werden beide in het onderzoek betrokken. Het onderzochte materiaal kan als volgt worden gerubriceerd:

A Kinderlectuur – algemeen	200.000 woorden
B Kinderlectuur – Indisch milieu	50.000 woorden
C Eenvoudige volwassenen lectuur – algemeen	600.000 woorden
D Eenvoudige volwassenen lectuur – Indisch milieu	150.000 woorden
	<hr/>
Totaal	1.000.000 woorden

Bij het onderzoek is rekening gehouden met het beginsel van de frequentie, de bruikbaarheid en de spreiding van de woorden. Onder frequentie verstaat men het aantal malen dat een woord in

\* Het Psychologisch Laboratorium van de Rijksuniversiteit te Utrecht heeft in 1963 'De la Court's frequentietelling van Nederlandse woorden' opnieuw laten verschijnen, als rapport nr. 6301 (V + 145 blz.). In deze uitgave zijn alle woorden opgenomen.

de teksten voorkomt.

De bruikbaarheid van een woord wordt bepaald door de mogelijkheid die het biedt voor de vorming van samenstellingen en afleidingen. Zo is bijvoorbeeld huis belangrijker dan buffer, omdat huis ook voorkomt in: woonhuis, stadhuis, gemeentehuis, huisdeur, huiskamer, huissleutel, enz.\*, terwijl het woord buffer o.a. voorkomt in bufferstaat, buffervoorraad. Komt een woord regelmatig voor in diverse teksten dan heeft dit woord een wijde spreiding. Komt het woord in een bepaalde tekst bijvoorbeeld 50 keer voor en in het geheel niet of zelden in een groot aantal andere teksten dan heeft dat woord een nauwe spreiding. Een woord met een ruime spreiding is meer algemeen bruikbaar en is dus van meer betekenis.

Vervolgens zijn woorden met dezelfde spelling maar met een verschillende betekenis opgenomen als verschillende woorden. De door De la Court en zijn medewerkers samengestelde woordfrequentielijst is het enige groots opgezette en wetenschappelijk verantwoorde onderzoek naar de woordfrequentie in de Nederlandse taal, voor zover ons bekend. In deze lijst zijn van de onderzochte 1.000.000 woorden de woorden opgenomen die een frequentie hebben van 25 of meer. Dat in deze lijst woorden als 'inlander, sloep en resident' wel genoemd zijn, terwijl bijvoorbeeld worden als 'kers, peer, pruim, en maaien' er niet in voorkomen, heeft men te aanvaarden.

De la Court heeft de 3296 woorden die 25 of meer keer zijn voorgekomen, in zijn rapport in alfabetische volgorde gepubliceerd.

Tevens heeft hij deze woorden vermeld, verdeeld over 7 radii, en wel:

1e radius (frequentie boven 500; 'ontelbaren')	305 woorden
2e radius (frequentie 300–500)	130 woorden
3e radius (frequentie 200–300)	169 woorden
4e radius (frequentie 150–200)	181 woorden

\*De *Woordenlijst Nederlandse Taal 1954* vermeldt reeds een honderdtal mogelijkheden met huis aan het begin van het woord.

5e radius (frequentie 100-150)	351 woorden
6e radius (frequentie 50-100)	878 woorden
7e radius (frequentie 25- 50)	1282 woorden
	<hr/>
	3296 woorden

4. *Aanvulling en interpretatie van de lijst De la Court voor dit onderzoek*

1. De lijst van de De la Court is voor ons doel aangevuld met:
  - a. De namen van de acht hoofdwindstreken (noord, oost, zuid, west, noordoost, noordwest, zuidoost, zuidwest).
  - b. De namen van maten en gewichten (centimeter, decimeter, hectometer, deciliter, liter, hectoliter, gram en ons)\*.
  - c. De namen van de dagen van de week (zondag, maandag, dinsdag, woensdag, donderdag, vrijdag en zaterdag).
  - d. De namen van de maanden (januari, februari, maart, april, mei, juni, juli, augustus, september, oktober, november en december).
  - e. De telwoorden en de rangtelwoorden.
  - f. De namen van landen (inclusief streken en provincies) volkeren en hun afleidingen.
2. Elk in de lijst opgenomen woord is inclusief de verbogen en vervoegde vorm(en). Bijvoorbeeld onder het werkwoord 'zijn' vallen: is, was, waren, geweest, enz. Ook de meervoudsvorm(en), de trappen van vergelijking en de verkleiningsvorm horen thuis bij het in de lijst opgenomen woord.  
Tevens werd bij het in de lijst opgenomen woord ondergebracht:
  - het vrouwelijk woord indien dit wordt gevormd door de gewone uitgang, bijvoorbeeld koning - koningin, eigenaar - eigenares;
  - een bijvoegelijk gebruikt woord dat bestond uit een in de lijst genoemd woord met de toevoeging van een 'e' of 'en';
  - het bijvoegelijk gebruikte verleden-deel-

\* Meter, kilometer, kilogram en pond zijn in De la Court's lijst vermeld.

woord van werkwoorden die in de lijst voorkomen;  
- het zelfstandig gebruikte verleden-deelwoord van werkwoorden die in de lijst voorkomen.

3. Komen samengestelde woorden zelf niet voor in de lijst, maar wel de afzonderlijke delen (in de betekenis die ze in de samenstelling hebben) dan werden ze beschouwd alsof ze in de lijst voorkomen, bijvoorbeeld goudgeel, ontdekkingsreiziger; niet woorden als bijziende, wittebroodsweken, onderweg.
4. De in de lijst opgenomen woorden zijn, indien geen nadere omschrijving is vermeld, in de meest ruime betekenis genomen.
5. De eigennamen\* en hun afleidingen zijn afzonderlijk geteld.

5. *Beschrijving van het onderzoek*

Bij dit onderzoek is van dezelfde indeling Nederlandse lectuur uitgegaan als bij de ontwikkeling van de formule voor de leesindex A (gemiddelde zins- en woordlengte)<sup>1</sup>.

De categorie studieboeken is echter moeten vervallen in verband met de vaktermen (milieuwoorden). Ook is 'Hop maar Jip en Janneke' van Annie M. G. Schmidt niet opgenomen.

De onderzochte boeken zijn:

I. a. *Kinderlectuur, leeftijd 6-7-8 jaar.*

1. Laan, Dick: Pinkeltje in Artis: Van Holkema en Warendorf N.V., Amsterdam.
2. Fransen, Frans: Puk en Muk op de tan-

\* Een eigenaam is een naam waardoor men een zelfstandigheid in het bijzonder een persoon maar bijvoorbeeld ook geografische zaken, van andere der zelfde soort onderscheidt, in tegenstelling met soortnaam (voorwerpsnaam die aan alle voorwerpen van dezelfde soort gemeen is).

dem; Drukkerij van het R.K. Jongensweeshuis, Tilburg.

b. *Kinderlectuur, leeftijd 9-10-11-12 jaar.*

3. Hildebrand, A. D.: Karel de kleinzoon van Daan de eend; H. Meulenhoff, Amsterdam.

4. Hölscher, To.: Hoe trotskopje genas; G. B. van Goor Zonen's Uitgeversmij. N.V., 's-Gravenhage-Djakarta.

c. *Kinderlectuur, leeftijd 13-14-15 jaar.*

5. Nowee, J.: Witte Veder; L. C. G. Malmberg, 's-Hertogenbosch.

6. Jansonius, Hella: In 'de Spreuwenpot'; U.M. 'West-Friesland', Hoorn.

## II. *Specifieke jeugdlectuur*

7. Steuben, Fritz: De rode storm; G. B. van Goor Zonen's Uitgeversmij. N.V., 's-Gravenhage-Djakarta.

8. Havelte, Sanne van: Onder de mimosa's; Van Holkema en Warendorf N.V., Amsterdam.

## III. a. *Detectives*

9. Christie, Agatha: Schuldig in eigen ogen; A.W. Sijthoff's Uitgeversmij. N.V., Leiden 1951.

10. Sayers, Dorothy L.: Onaangenaamheden in de Bellona club; Uitgeverij 'Pax', 's-Gravenhage.

b. *Cowboys*

11. Brand, Max: Tijger; Helicon N.V. Uitgeversmij., Antwerpen.

12. Macdonald, Wm Colt: 'Hands up holiday', Uitgeversmaatschappij 'De Combinatie' N.V., Rotterdam 1952.

c. *Dames-romans*

18. Elzer, Margarete: Effy zet door! Zuid-Hollandsche Uitgeversmij., 's-Gravenhage.

14. Deeping, Warwick: Levensernst; Uitgeverij 'De Steenuil', Amsterdam.

## IV. *Reisverhalen*

15. Juliën, Dr. Paul: Kampvuren langs de evenaar; Poirtersfonds-serie, Eindhoven.

16. Pos, Mary: Op vleugels naar Zuid-Afrika; C.V. Allert de Lange, Amsterdam 1941.

b. *Historische romans*

17. Haasse, Hella S.: Het woud der verwachting; N.V. Em. Querido's Uitgeversmaatschappij, Amsterdam 1949.

18. Ammers-Küller, Jo v.: De Patriotten I; Engelhard, Van Embden en Co., Amsterdam.

c. *Streekromans*

19. Man, Herman de: Een stoombootje in de mist; De Nederlandsche Keurboekerij, Amsterdam 1933.

20. Koenen, Marie: Wassend graan; Uitgeverij Het Spectrum, Utrecht-Brussel.

d. *Populair-wetenschappelijke boeken*

21. Zischka, Anton: Europa leeft; Nederlands Boekhuis, Tilburg.

22. Edelman, Mr. H. C. M.; Leven met atomen; J. M. Meulenhoff, Amsterdam.

De boeken zijn doorgemeten met behulp van steekproeven. De bepaling van de grootte van de steekproef per boek is als volgt gebeurd.

In het geval, dat de woorden uit de aangevulde lijst-De-la-Court (geïnterpreteerd zoals in het voorgaande is beschreven) een percentage vertegenwoordigen van respectievelijk 90, 80 of 70, dan is - bij een steekproefgrootte van 2500 woorden per boek - de maximale fout in dit percentage respectievelijk gelijk aan plus of min 1,2%, 1,6% of 1,8% behoudens een kans van 5%. Daar uit het verslag van het 'Onderzoek De la Court' niet valt op te maken welk percentage de vermelde woorden in werkelijkheid vertegenwoordigen,

werd voor een tweetal van de door hem onderzochte teksten het percentage berekend. Dit gaf aanleiding om aan te nemen dat door de woorden uit deze lijst minstens 80 en misschien zelfs 90% van de onderzochte teksten wordt vertegenwoordigd.

Bij aanneming van 80% geeft een steekproef van 2500 woorden dus een fout van plus of min 1,6% behoudens een kans van 5%. Gebleken is dat de lijst De la Court met aanvulling in ons onderzoek 92,5% van alle onderzochte teksten

bevat.\*

Wij namen per boek negen willekeurig gekozen passages van  $\pm 300$  woorden, zodat de totale steekproef per boek bestond uit  $\pm 2700$  woorden. Na aftrek van de eigennamen en hun afleidingen bleven steeds meer dan 2500 te onderzoeken woorden over.

Van alle woorden uit de verschillende steekproeven is nagegaan of zij in de aangevulde lijst De la Court voorkwamen. (De zogenaamde ontelbaren zijn ook apart geteld).

Tabel I Overzicht per schrijver van percentage outsiders.

Schrijver	percentage outsiders	percentage verschillende outsiders	indexcijfer B	percentage gewogen outsiders		
				gemiddeld	minimum behoudens een kans van 5%	maximum
1 Dick Laan	5,114	2,217	85	2,870	2,403	3,337
2 Frans Franssen	3,629	2,603	85	2,887	2,317	3,475
3 A. D. Hildebrand	2,813	1,771	91	2,105	1,644	2,566
4 To Hölscher	3,939	3,246	80	3,433	2,785	4,081
5 L. Nowee	5,724	4,356	71	4,666	3,948	5,384
6 Hella Jansonius	7,415	5,889	56	6,389	5,528	7,250
7 Fritz Steuben	6,864	5,577	59	6,038	5,197	6,879
8 Sanne v. Havelte	5,948	4,679	66	5,164	4,382	5,946
9 Agatha Christie	4,665	3,899	74	4,159	3,441	4,877
10 Dorothy L. Sayers	7,195	5,929	57	6,336	5,471	7,201
11 Max Brand	5,851	5,393	64	5,534	4,701	6,367
12 Wm Colt Macdonald	6,811	5,346	61	5,758	4,943	6,573
13 Margarete Elzer	7,190	5,985	56	6,478	5,589	7,367
14 Warwick Deeping	5,844	4,942	66	5,195	4,403	5,987
15 Dr. Paul Juliën	10,163	8,391	36	8,992	7,975	10,009
16 Mary Pos	7,555	6,793	51	7,092	6,146	8,038
17 Hella S. Haasse	10,706	9,617	28	9,982	8,902	11,062
18 Jo v. Ammers-Küller	9,834	8,934	33	9,349	8,295	10,403
19 Herman de Man	7,917	6,729	51	7,125	6,203	8,047
20 Marie Koenen	7,308	6,608	52	6,879	5,963	7,795
21 Anton Zischka	9,547	8,163	38	8,717	7,719	9,715
22 Mr. H. C. M. Edelman	14,891	9,257	22	10,688	9,720	11,656

\* Ons onderzoek omvatte 61 391 woorden, waarvan 2 632 woorden eigenamen of afleidingen daarvan waren. Van de overgebleven 58 759 woorden bleek 71,5% te behoren tot de 300 'ontelbaren' van De la Court. De resterende 3000 woorden van de aangevulde lijst van De la Court bestreken nog eens 21%. Men kan dus stellen dat men met behulp van de

woordfrequentielijst De la Court (bestaande uit 3300 woorden) 92,5% van de door ons doorgemeten teksten kan begrijpen.

Wij menen hieruit te mogen concluderen dat de 'Lijst De la Court' ontstaan in 1934 heden nog zeer bruikbaar is.

Een niet in de lijst voorkomende woord is een 'outsider'. Van deze outsiders is tevens de frequentie nagegaan.

In dit onderzoek is aangenomen dat wanneer een outsider éénmaal voorkomt, deze outsider voor de lezer zijn volle moeilijkheid bezit en de waarde heeft van 1. Komt een outsider tweemaal voor dan is de tweede keer de moeilijkheid gedaald tot de helft (deze outsider krijgt dan de waarde 1,5). Komt een outsider driemaal voor dan is de derde keer de waarde 0,25 en wordt deze outsider geclassificeerd met een waarde van 1,75. De outsiders met een frequentie van 4 worden bij de vierde maal geteld als 0,125. Als een outsider meer dan viermaal in de tekst voorkomt, wordt verondersteld dat de lezer de vijfde en volgende keren de betekenis ervan kent en dat dus dit woord voor hem geen moeilijkheden meer oplevert. Komt een outsider bijvoorbeeld achttien keer in de tekst voor, dan wordt deze outsider herleid tot een waarde van 1,875 outsider. De herleide outsiders en de eenmaal voorkomende outsiders vormen samen de 'gewogen' outsiders.

Nadat het aantal gewogen outsiders per boek was vastgesteld, werden de percentages gewogen outsiders per boek berekend.\*

\* Het percentage gewogen outsiders is bepaald t.o.v. de totale steekproef per boek, dus inclusief de eigenamen.

Om de nauwkeurigheid te kunnen aangeven zijn 95%-betrouwbaarheidsgrenzen voor de percentages bepaald. In Tabel 1 ziet men een en ander per boek/schrijver weergegeven.

De percentages gewogen outsiders zijn uitgedrukt in indexcijfers B (indexcijfer A is gereserveerd voor de leesbaarheidsfactor: gemiddelde zinslengte, gemiddelde woordlengte en hun onderlinge betrekking). De spreidingsband van het indexcijfer B loopt, eveneens als bij indexcijfer A, van 1 t/m 100, met dien verstande dat 1% gewogen outsiders overeenkomt met het indexcijfer B = 100 en 13,5% gewogen outsiders met het indexcijfer B = 0. Voorlopig is een lineaire verdeling aangehouden.

De verhouding van het percentage gewogen outsiders en de leesindex B is weer te geven in de formule  $y = -8x + 108$  (waarin  $y$  = leesindexcijfer B en  $x$  = percentage gewogen outsiders).

Door samenvoeging van de verschillende typen lectuur en door afronding ontstaat het in Tabel 2 gegeven overzicht.\*

\* Een uitzondering hierop vormt Dr. Paul Juliën met zijn *'Kampvuren langs de evenaar'*. De moeilijkheidsgraad van dit boek gezien vanuit de factor bekendheid van de woorden ligt op het peil van de historische romans en populair-wetenschappelijke lectuur. Naar de inhoud zou plaatsing van dit boek in het genre populair-wetenschappelijk alleszins verantwoord zijn.

Tabel 2 Leesbaarheid bepaald door de factor woordkeus.

Categorie	Indexcijfer B	Omschrijving	Type lectuur
I	76-100	gemakkelijk	kinderlectuur t.m. 12 jaar
II	56- 80	standaard	kinderlectuur vanaf 13 jaar, specifieke jeugdlectuur, detectives, cowboys en damesromans
III	36- 60	tamelijk moeilijk	reisverhalen en streekromans
IV	16- 40	moeilijk	historische romans en populair-wetenschappelijke lectuur
V	0- 20	zeer moeilijk	wetenschappelijke lectuur

Literatuur

1. Brouwer R. H. M.: *Onderzoek naar de leesbaarheid van Nederlands proza; Pedagogische Studiën*, jrg. 40 nr. 10 (okt. 63).
2. Court, A. De la e.a.: *Lijst van de in het Nederlandsch meest voorkomende en meest bruikbare woorden. Verslag van een onderzoek in opdracht van het Departement van Onderwijs en Eeredienst; Uitgegeven bij Volkslectuur te Batavia, 1937.*

Curriculum Vitae

R. H. M. Brouwer, geb. in 1929 te Heerlen. Behaalde

in 1967 de akte Pedagogiek MO B. Is werkzaam geweest bij de Staatsmijnen in Limburg bij de afdeling Opleiding en Kadervorming en de afdeling Personeelsresearch. Sedert enige jaren verbonden aan het Onderwijs-Adviescentrum te Amersfoort. Publiceerde 'Enige aspecten van het beter en sneller lezen' (Richting 1959/1960); 'Leesbaarheid van Nederlands proza' (Pedagogische Studiën 1963); 'Geprogrammeerde cursus: Bediening pneumatische regelaar Foxboro M 54' (in samenwerking met Dr. G. Koene en T. Vandermuehlen in Geprogrammeerde Instructie 1966); 'Taalonderwijs aan buitenlandse werknemers' (Mens en Onderneming 1966); 'Verstaanbaarheid en meetbaarheid van de taal' (Dux 1966).

Indicatie	Indicatie 2	Omschrijving	Type
1	75-100	...	...
2	25-80	...	...
3	10-50	...	...
4	0-20	...	...