

## Hernieuwde "Streeflijst Woordenschat voor 6-jarigen, gebaseerd op onderzoek in Nederland en België": een tussentijds onderzoeksverslag

M. Lejaegere en A. Schaerlaekens

*Universitair Ziekenhuis St.-Rafaël, Leuven*

In 1981 verscheen de "Nieuwe Streeflijst Woordenschat voor 6-jarigen, gebaseerd op onderzoek in Nederland en België", als resultaat van een gemeenschappelijk onderzoeksprogramma van de Rijksuniversiteit Leiden en de Katholieke Universiteit Leuven. In deze streeflijst werd de woordenschat samengebracht waarvan de relatieve mate van passieve beheersing door 6-jarigen met unanimité wordt verondersteld door ervaren leerkrachten bij aanvang van het lager onderwijs. Het huidige onderzoek is gericht op het uitbrengen van een derde versie van de Streeflijst Woordenschat voor 6-jarigen. In dit onderzoeksverslag wordt de onderzoeksmethodiek beschreven naar samenstelling van de nieuwe uitganglijst en de verwerkingsmogelijkheden.

### **Inleiding**

Het FWO-project (Fonds voor Wetenschappelijk Onderzoek - Vlaanderen (het vroegere N.F.W.O., Nationaal Fonds voor Wetenschappelijk Onderzoek)) dat momenteel loopt op het Centrum voor Taalverwervingsonderzoek te Leuven, omvat de actualisering van de "Nieuwe Streeflijst Woordenschat voor 6-jarigen, gebaseerd op onderzoek in Nederland en België" (Kohnstamm, Schaerlaekens, De Vries, Akkerhuis, & Frooninckx, 1981). In 1971 kwam de eerste versie van deze streeflijst tot stand, die in 1981 bewerkt werd tot de tweede versie. Actualisering van een dergelijke lijst is vanwege de sterke onderhevigheid aan veroudering een permanente opdracht. Het huidige onderzoek zal bijgevolg resulteren in een derde uitgave van dit woordenschatonderzoek.

De methodologie van het huidige onderzoek is gelijkaardig aan die van de vorige versies, namelijk passieve woordenschat via indirecte methode nagaan.

---

Correspondentie-adres: M. Lejaegere, Dienst Neus-, Keel-, Oorzakten, Gelaat en Halschirurgie, Centrum voor Taalverwervingsonderzoek, UZ St.-Rafaël, Kapucijnenvoer 33, 3000 Leuven, België. Tel. 016/33.23.52. Fax. 016/33.23.27. E-mail: maryline.lejaegere@uz.kuleuven.ac.be

Concreet betekent dit dat van een aantal woorden wordt nagegaan hoeveel kleuterleid(st)ers (derde kleuterklas in België, groep 2 in Nederland) en leerkrachten van het lager onderwijs (eerste leerjaar in België, groep 3 in Nederland) vinden dat die woorden begrepen zouden moeten worden, als zij zouden voorkomen in een eenvoudige, gesproken zin. De woorden moeten dus niet actief gebruikt worden door de kinderen, wel passief begrepen worden. Ook hoeft het feitelijke begrip van de woorden bij de kinderen niet nagegaan te worden, wel of hun leerkrachten vinden dat ze die woorden zouden moeten begrijpen. Vandaar ook de benaming 'streeflijst', doelend op de verwachtingen van de leerkrachten in verband met de woordenschat van hun leerlingen.

### **Onderzoeksmethode**

De hernieuwing van de Streeflijst Woordenschat startte officieel op 1 januari 1995. Het is opnieuw een samenwerkingsprogramma tussen Nederland en België. De werkwijze van het project is op te delen in verschillende stappen.

#### *Samenstellen van de woordenschatbronnen*

Na een grondige literatuurstudie van het reeds verrichte Nederlandstalig en vergelijkbaar buitenlands onderzoek op het vlak van woordenschat en woordenschatlijsten, werd een uitgebreid vooronderzoek gestart. Dit had als doel een aantal woordenschatbronnen samen te stellen om als basis te fungeren voor de nieuwe uitgangslijst, die aan de beoordelaars uit het onderwijs zal worden voorgelegd. In totaal werden er een zestal Noord- en Zuid-Nederlandse bronnen geraadpleegd, bewerkt of samengesteld.

Voor de Noord-Nederlandse bronnen werd er enerzijds beroep gedaan op bestaand woordenschatonderzoek, meer bepaald het onderzoek van Coenen en Vermeer naar de Nederlandse woordenschat van allochtone kinderen (1988) en Schrooten en Vermeer naar de Nederlandse woordenschat in het basisonderwijs (1994). Uit deze bronnen werd enkel de met de Streeflijst leeftijdsovereenkomstige woordenschat geselecteerd, met name de woordenschat van 5- à 6-jarigen. Anderzijds werd er een eigen onderzoek opgestart naar de woordenschat van het kindertelevisieprogramma *Sesamstraat* (Depré, 1996). De "Nederlandse *Sesamstraat*" bestaat grotendeels uit hetzelfde beeldmateriaal als de Amerikaanse versie, maar dan met een Nederlandse vertaling ( $\pm 10$  minuten zendtijd). De begin- en eindscène van elke aflevering daarentegen zijn origineel Nederlands materiaal (samen  $\pm 3$  à  $4$  minuten zendtijd). Voor dit onderzoek werden de draaiboeken of scripts van de Nederlandse onderdelen (begin- en eindscènes) van 1 jaar *Sesamstraat* (september 1994–september 1995) op computer ingebracht. Concreet gaat het hier om 149 verschillende scènes. Aangezien de scripts geen schriftelijke neerslag van volledige *Sesamstraat*afleveringen zijn, hebben we aanvullend nog 47 verschillende afleveringen (totaal 11 uur 45 minuten) opgenomen van televisie, uitgeschreven en ingebracht op computer. Eventueel overlappende

stukken tussen de scripts en de afleveringen werden weggelaten. Na het intikken van alle Sesamstraatgegevens, voerde de computer aan de hand van het programma WORDFREQ. (Word for DOS) een woordfrequentieberekening door, met als resultaat een alfabetische lijst met een frequentiegetal per verschillend woord. De Sesamstraatscripts leverden totaal 3.886 verschillende woorden op, de afleveringen 5.435 woorden. Samen resulteerde dit in een overkoepelende lijst van 7.246 woorden. Na extractie van alle klanknabootsingen, eigennamen, nonsenswoorden, e.a. resteerden in de overkoepelende lijst nog 6.462 verschillende woorden of types. Alle types werden nadien voorzien van een gedetailleerd betekenisonderscheid, dat werd toegekend volgens woordsoort en op basis van Van Dale's Groot Woordenboek der Nederlandse Taal.

Voor het Zuid-Nederlandse luik werd er ook beroep gedaan op bestaande bronnen. Dit betrof het onderzoek van Jaspaert en Schrooten naar de woordenschat in het Vlaamse onderwijs (in press), aangevuld met een eigen onderzoek naar de rekenwoordenschat (De Cuyper, 1994) en een eigen woordfrequentieonderzoek op basis van 22 taalsamples van kinderen. Uit het onderzoek van Jaspaert en Schrooten werd, analoog met de Noord-Nederlandse bronnen, enkel de met de Streeflijst leeftijdsovereenkomstige woordenschat geselecteerd. Wat het onderzoek naar de rekenwoordenschat betreft, baseerden we ons op het onderzoek van De Cuyper (1994), die de woordenschat van vier van de in Vlaanderen veelgebruikte rekenmethoden in het eerste leerjaar in kaart bracht. Alleen de unanieme woorden voor alle vier rekenmethoden werden geselecteerd en integraal opgenomen in de nieuwe uitgangslijst. Als derde Zuid-Nederlandse bron hebben we 22 uitgeschreven observaties van gemiddeld 4 uur spontane taal van kinderen tussen 4;1 en 5;10 jaar ingebracht op computer. Opnieuw werd een woordfrequentieberekening doorgevoerd aan de hand van het programma WORDFREQ. (Word for DOS), hetgeen resulteerde in een lijst van 5.017 woorden. Na extractie van klanknabootsingen, eigennamen, e.a. bleef er nog een lijst van 3.937 types over. Alle types werden voorzien van een gedetailleerd betekenisonderscheid, op dezelfde manier als bij het Sesamstraatonderzoek.

### *Samenstellen van de nieuwe uitgangslijst*

De nieuwe uitgangslijst werd tot stand gebracht in twee stappen, namelijk ten eerste een actualisering van de vorige versie, de Streeflijst Woordenschat 1981, op basis van de Noord- en Zuid-Nederlandse bronnen en ten tweede, de toevoeging van nieuwe, recente woorden (anno 1995-1997).

Voor het eigenlijke actualiseren werd de Streeflijst 1981 gesplitst in twee delen: de woorden met een frequentie lager dan of gelijk aan 40% en de woorden met een frequentie hoger dan 40%. De woorden met een frequentie lager dan of gelijk aan 40% werden voorgelegd aan laatstejaarsstudenten van het Pedagogisch Hoger Onderwijs – afdeling Kleuteronderwijs in Nederland (Arnhem) en in België (Heverlee). In een eerste fase werden 1.099 woorden met een frequentie tussen 20% en 40% vermengd met afleiders (hoogfrequente(re) woorden) en voorgelegd aan 50 studenten (30 voor België en 20 voor Nederland). Zij kregen

dezelfde instructie als de beoordelaars van het onderzoek van 1981, tevens ook de instructie voor de beoordelaars van het huidig onderzoek. Deze instructie luidt: "Vind ik dat deze woorden, door 6-jarigen die ik ken, begrepen moeten worden, indien deze woorden zouden voorkomen in een eenvoudige gesproken zin?" Alle woorden waarvan meer dan de helft van de beoordelaars vonden dat ze niet begrepen moeten worden door 6-jarigen, werden geschrapt. Het resterende deel omvatte 621 woorden. In een tweede fase werden die 621 woorden samengevoegd met de woorden van de Streeflijst 1981 met een frequentie lager dan 20% (totaal 534 woorden), en vormden op die manier een lijst woorden met een frequentie lager dan of gelijk aan 40%. Deze lijst bedroeg in totaal 1.155 woorden en werd opnieuw vermengd met afleiders (hoogfrequente(re) woorden) en voorgelegd aan 61 studenten (34 voor België en 27 voor Nederland), met dezelfde instructie als voorheen. Opnieuw werden alle woorden, waarvan meer dan de helft van de beoordelaars vonden dat ze niet begrepen moeten worden door 6-jarigen, geschrapt. Dit resulteerde in een lijst van 503 woorden die heropgenomen worden in de uitgangslijst van de Hernieuwde Streeflijst Woordenschat (1997).

De woorden van de Streeflijst 1981 met een frequentie hoger dan 40% (totaal 5.152 woorden) werden systematisch 'gematched' aan de woorden van de 5 Noord- en Zuid-Nederlandse bronnen. Aangezien de woordenschat uit de bron Rekenwoordenschat quasi uniek is voor deze bron en weinig of niet terug te vinden is in de andere bronnen, werd hij apart behandeld. Alle unanieme woorden voor de vier rekenmethoden werden, in tegenstelling tot de woordenschat van de andere 5 bronnen, integraal opgenomen in de uitgangslijst. Dit betekent dat er 854 rekenwoordenschatwoorden toegevoegd werden aan de uitgangslijst, zonder ze te matchen aan de woorden van de andere bronnen. Bij het matchen van de 5.152 woorden van de Streeflijst 1981 aan de woorden van de 5 overige bronnen, werd het volgende criterium gehanteerd: indien het woord van de Streeflijst 1981 in minstens 1 van de 5 bronnen voorkomt, dan wordt het woord opgenomen in de nieuwe uitgangslijst. Alle woorden die qua betekenis verwarrend of overlappend waren, werden voorzien van één of meerdere gedetailleerde betekenisonderscheiden. Dit resulteerde in een lijst van 4.764 woorden.

Het overblijvend woordmateriaal van de 5 bronnen na het matchen, vormde de basis voor het toevoegen van nieuwe, recente woorden (anno 1995–1997). Aangezien er geen recent objectief uitgangspunt is waarmee het overblijvend woordmateriaal kon worden vergeleken, werden alle woorden beoordeeld en eventueel toegevoegd op basis van de unanieme subjectieve indrukken van 2 logopedisten, vertrouwd met het taalgebruik van 5- à 6-jarigen. Per bron betekent dit het volgende aantal toegevoegde, nieuwe of recente woorden: bron Jaspaert en Schrooten 68 woorden, bron taalsamplesonderzoek 64 woorden, bron Coenen en Vermeer 90 woorden, bron Sesamstraatonderzoek 57 woorden en bron Schrooten en Vermeer 154 woorden. Tenslotte werden nog 20 woorden toegevoegd als eigen suggesties, zoals beeldscherm (t.v., computer), cd, dinosaurus, shoppingcentrum, parfum, babyfoon, droogkast, printer ... Dit alles vormt een lijst van

6.574 woorden, die onderling overlappend zijn. Na extractie van de dubbele woorden, bleef er een lijst over van 5.689 woorden. In vergelijking met de uitgangslijst van 1981 (totaal 6.785 woorden), bevat deze lijst van 5.689 woorden weinig woorden die vermoedelijk laagfrequent zullen scoren. Deze woorden waren blijkbaar iets te goed uitgezuiverd door de laatstejaarsstudenten Pedagogisch Hoger Onderwijs (cf. hoger). Om te vermijden dat er geconcludeerd moet worden dat de huidige kleuters minder woorden begrijpen dan de kleuters van 15 jaar geleden, hebben we de lijst van 5.689 woorden aangevuld met 691 laagfrequente(re) woorden van de Streeflijst 1981. Samen vormen ze de nieuwe uitgangslijst van 6.380 woorden. Het samenstellen van de uitgangslijst en het verwerken van de bronnen gebeurde computergestuurd aan de hand van EXCEL 5.0 for Windows en ACCESS 2.0 for Windows.

### ***Beoordeling van de lijst door leerkrachten***

Deze uitgangslijst wordt beoordeeld door 160 beoordelaars, 80 in België en 80 in Nederland. Per land komt dit theoretisch neer op 40 beoordelaars uit het kleuteronderwijs (derde kleuterklas-groep 2) en 40 beoordelaars uit het eerste leerjaar/groep 3. Deze beoordelaars zijn verkregen via een gestratificeerde steekproef, aselekt getrokken, met als stratificatiecriteria per land: onderwijsnet, provincies en onderwijsgraad (kleuter- of lager onderwijs). Deze beoordelaars werden op voorhand gecontacteerd om hun medewerking te vragen. In een eerste steekproef werden 120 scholen gecontacteerd in België en 108 in Nederland. Hiervan verleenden 67 scholen in België hun medewerking aan het onderzoek en 47 scholen in Nederland. Via een tweede aanvullende steekproef, respectievelijk 30 scholen in België en 111 scholen in Nederland, werden deze aantallen opgetrokken tot 75 deelnemende scholen voor België (35 beoordelaars derde kleuterklas, 40 beoordelaars eerste leerjaar) en 86 deelnemende scholen voor Nederland (46 beoordelaars groep 2, 40 beoordelaars groep 3).

De uitgangslijst werd gedrukt op optisch leesbare formulieren (OMR-papier), zodat de gegevens computergestuurd gelezen en verwerkt kunnen worden. Elke beoordelaar kreeg een dergelijke lijst toegestuurd en scoorde elk woord waarvan hij/zij vindt dat 6-jarigen dat woord moeten begrijpen. Het scoren zelf bestaat uit het zwart maken of kleuren van het blanco hokje vóór elk woord. Nadien stuurden de beoordelaars de ingevulde lijsten terug.

### ***Verwerking van de ingevulde lijsten***

Elke lijst werd per bladzijde ingescand met behulp van een optische lezer, die de scoring omzet in een ASCII-gegevensbestand. Dit bestand wordt vervolgens ingelezen in ACCESS for Windows en omgezet in beoordelingspercentages. De verwerkte resultaten zullen weergegeven worden in een aantal lijsten, namelijk

- \* een alfabetische woordenlijst met per woord waarderingsgetallen, opgesplitst volgens waarderingsgetal per land en een globaal waarderingsgetal
- \* een unaniemenlijst (minstens 90% van de beoordelaars oordeelt dat het woord op 6-jarige leeftijd begrepen moet worden)

- \* lijsten van woorden met lagere frequenties, opgesplitst volgens frequentieregio's (bv. woorden met waarderingsgetal tussen 75% en 90%)
- \* een lijst woorden met grote verschillen in waarderingsgetal tussen Nederland en België (minstens 35% verschil)
- \* een lijst termen die onontbeerlijk zijn voor beginnend rekenonderricht.

### **Analyse van de resultaten**

Tenslotte zullen de gegevens statistisch geanalyseerd worden naar o.a. verschillen in beoordeling tussen Nederlandse en Belgische beoordelaars, tussen mannelijke en vrouwelijke beoordelaars, tussen beoordelaars uit het kleuteronderwijs en beoordelaars uit het lager onderwijs en verschillen in beoordeling naargelang het aantal jaren ervaring in het onderwijs. Ook een vergelijking tussen 1981 en 1997, wat betreft de evolutie van de Noord-Zuid verschillen, de evolutie van de dialectgekleurde woorden, en de inhoudelijke veranderingen verbonden met de verschuivende interesse in de leefwereld van 6-jarigen, zal op basis van de resultaten uitgewerkt worden.

### **Tot slot ... enkele streefdata**

De ingevulde lijsten zijn reeds ingescand en beschikbaar op computer. Momenteel werken we aan de berekening van de beoordelingspercentages en de statistische analyse van de gegevens. Het samenstellen van de uiteindelijke frequentielijst wordt vermoedelijk afgerond eind 1997/begin 1998.

### **Summary**

In 1981, a vocabulary list for six-year-olds was published, resulting from a common research program of the University Leiden (the Netherlands) and the Catholic University Leuven (Belgium). This comprehensive vocabulary list was based on judgements of kindergarden teachers and teachers of the first grade. The current research project, discussed in this paper, will result in the third version of this vocabulary list.

### **Literatuur**

- Coenen, M., & Vermeer, A. (1988). *Nederlandse woordenschat allochtone kinderen*. Tilburg: Zwijsen.
- De Cuyper, K. (1994). *De woordenschat in rekenmethoden. Literatuurstudie en opstellen van een woordenschatlijst voor het eerste leerjaar vanuit vier van de in Vlaanderen veelgebruikte rekenmethoden*. K.U. Leuven, niet-gepubliceerde eindverhandeling, faculteit Geneeskunde, afdeling Logopedie en Audiologie.
- Depré, B. (1996). *Televisie, kinderen en woordenschat. Inclusief een praktijkgericht onderzoek naar de woordenschat van het kindertelevisieprogramma "Sesamstraat" in functie van de hernieuwde streeflijst*. K.U. Leuven, niet-gepubliceerde eindverhandeling, faculteit Geneeskunde, afdeling Logopedie en Audiologie.

- Jaspaert, K., & Schrooten, W. (in press, vermoedelijk eind 1998) *De school aan het woord. De taalvaardigheidseisen in het Vlaamse onderwijs.*
- Kohnstamm, G.A., & de Vries, A.K. (1971). *Streeflijst Woordenschat bij de overgang van kleuteronderwijs naar basisonderwijs.* IJmuiden: Vermande Zonen.
- Kohnstamm, G.A., Schaerlaekens, A., de Vries, A.K., Akkerhuis, G.W., & Frooninckx, M. (1981). *Nieuwe Streeflijst Woordenschat voor 6-jarigen, gebaseerd op onderzoek in Nederland en België.* Lisse: Swets & Zeitlinger.
- Schrooten, W., & Vermeer, A. (1994). *Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen.* Tilburg, Tilburg University Press.
- Van Dale (1992). *Groot Woordenboek der Nederlandse Taal.* Utrecht/Antwerpen, Van Dale Lexicografie.