

## **De school aan het woord.**

### **Een onderzoek naar de lexicale samenstelling van het taalaanbod op de basisschool in Vlaanderen en Nederland**

#### **1. Inleiding**

Dat een aantal leerlingen in het basisonderwijs niet taalvaardig genoeg is om in de klas te kunnen functioneren is al vaker vastgesteld. Om deze leerlingen de nodige taalvaardigheid bij te brengen is in de eerste plaats een goede beschrijving nodig van de behoeften waaraan het taalvaardigheidsonderwijs tegemoet moet komen. Deze behoeften zijn direct verbonden met het talige domein, het geheel van verwante taalgebruikssituaties waarin leerlingen moeten functioneren. In deze bijdrage bekijken we het domein van het 'schoolse leren'.

Tussen 1991 en 1996 werd op het Steunpunt NT2 van de Katholieke Universiteit Leuven onderzoek gedaan naar de toegankelijkheid van het taalaanbod op school. Er werden data verzameld in Vlaanderen en Nederland bij verschillende doelgroepen: basisonderwijs, voortgezet en buitengewoon onderwijs. In wat volgt beperken we ons tot de onderzoeksresultaten die betrekking hebben op het gewoon en buitengewoon basisonderwijs.

De 'onderwijstaal' kan op verschillende niveaus beschreven worden. Om te beginnen kunnen we aangeven welke typische talige taken leerlingen moeten kunnen uitvoeren, willen ze met enige kans op succes aan het onderwijs deelnemen. Het gaat dan om taken als: luisteren naar instructies van de leerkracht, leerstof verwerken op basis van zaakvakteksten, deelnemen aan een groepsgesprek, informatie halen uit naslagwerken etc. Deze taken vereisen het uitvoeren van bepaalde taalhandelingen en de beheersing van een aantal taalelementen en deelvaardigheden. Om op een degelijke manier te kunnen deelnemen aan een groepsgesprek moeten leerlingen bijvoorbeeld gericht kunnen luisteren, hoofdzaken en bijzaken kunnen onderscheiden en over een bepaalde woordenschat beschikken.

Het belang van deze laatste factor is erg groot. Een groot deel van de betekenis van het 'schoolse taalaanbod' wordt gedragen door de woordenschat waaruit het is samengesteld. In wat volgt, bekijken we het verzamelde taalaanbod dan ook in de eerste plaats vanuit dat perspectief, zonder evenwel uit het oog te verliezen dat het toegang krijgen tot het taalgebruik op school niet uitsluitend afhangt van kennis van woorden.

De beschrijving van de lexicale taalvaardigheidseisen gebeurde in verschillende stappen. Eerst werd een woordenlijst samengesteld die representatief moest zijn voor het vakoverschrijdende taalaanbod op de basisschool. De representativiteit van de woordenlijst werd in de eerste plaats verzekerd door woorden te selecteren op basis van de analyse van een representatief corpus bestaande uit mondeling taalaanbod van de leerkracht in de klas en schriftelijk taalaanbod uit schoolboeken. De samenstelling

van dit corpus wordt beschreven in paragraaf 2.1. Van dit corpus werd een frequentielijst gemaakt waaruit dan weer die woorden geselecteerd moesten worden die behoren tot de 'vakoverstijgende schooltaal'. Dit betekent dat een aantal categorieën van woorden uit de volledige frequentielijst verwijderd moesten worden: woorden die toevallig voorkomen en woorden die behoren tot de typische vaktaal van één enkel schoolvak. Om de eerste categorie woorden uit te sluiten werd een frequentie criterium gehanteerd: woorden die niet boven een bepaalde frequentiegrens uitkwamen, werden verwijderd. De tweede categorie werd uitgesloten door bij de selectie rekening te houden met de spreiding van een woord over verschillende corpusonderdelen. Wat de omvang van de lijst betreft, lieten we ons leiden door de overweging dat een leerling die de woorden uit de woordenlijst beheerst, in staat zou moeten zijn om met een willekeurige schooltekst om te gaan. We streven met andere woorden naar een voldoende hoge tekstdekking van de lijst ten opzichte van het volledige corpus. We gaan dieper in op de samenstelling van de woordenlijst in paragraaf 2.2.

In tweede instantie werd de woordenlijst gebruikt om een indicatie te geven van de moeilijkheidsgraad en de toegankelijkheid van bepaalde corpusonderdelen. Het criterium dat we daarvoor gebruiken, is de tekstdekking die we met de woordenlijst halen voor die corpusonderdelen. De woordenlijst is weliswaar samengesteld met de bedoeling een zo groot mogelijk gedeelte van het schoolse taalaanbod te dekken, maar dit impliceert geenszins dat we daarom ook voor elk corpusonderdeel voldoende hoge dekkingspercentages halen. Het is immers perfect mogelijk dat we met onze selectiecriteria zoveel woorden uitsluiten dat we toch niet aan de beoogde tekstdekking komen. Omgekeerd is het ook mogelijk dat onze selectiecriteria zo soepel zijn dat er te veel woorden in de lijst belanden - 'te veel' is hier dan vooral te begrijpen vanuit het standpunt van diegenen die de lijst als basis voor de selectie van leerdoelen zouden willen hanteren. In verband met de toegankelijkheid van de verzamelde data werden ook verschillen tussen leerjaren en tussen mondeling en schriftelijk materiaal onderzocht. De resultaten van deze tweede onderzoeksfase worden beschreven in paragraaf 2.3.

In een derde fase werd de woordenlijst door een vergelijking met andere representatieve corpora gevalideerd als beschrijving van 'algemene vakoverschrijdende schooltaal'. We berekenden hiervoor de dekkingspercentages over een corpus uit het Vlaamse speciaal onderwijs en een Nederlands corpus uit het reguliere basisonderwijs. Ook voor deze corpora keken we naar eventuele verschillen tussen corpusonderdelen. Dit deel van het onderzoek wordt beschreven in paragraaf 3.

De woordenlijst die in dit onderzoek werd samengesteld en gebruikt, heeft in de eerste plaats een functie als ijkpunt waartegen de lexicale complexiteit van de verschillende corpusonderdelen afgezet kan worden. Ze kan op dezelfde manier gebruikt worden bij het ontwikkelen van lesmaterialen. Het was uitdrukkelijk niet de bedoeling om te komen tot een soort streefwoordenlijst van 'woorden die leerlingen zeker moeten kennen'. Het formuleren van onderwijsdoelstellingen is een volgende stap waarbij de resultaten van dit onderzoek eventueel nuttig kunnen zijn. In

paragraaf 4 gaan we iets dieper in op de bruikbaarheid van onze onderzoeksresultaten voor de onderwijspraktijk.

## 2. Schooltaal in Vlaanderen

### 2.1. Samenstelling van het Vlaamse corpus

Op achttien verschillende Vlaamse basisscholen werd tussen januari en juni 1991 mondeling taalaanbod van leerkrachten verzameld. In elke klas werden gedurende een halve dag (voor- of namiddag) bandopnames gemaakt van de leerkrachtentaal. Op die manier werd een corpus van 686.035 *tokens* aangelegd. Een *token* is de concrete verschijningsvorm van een woord (*type*) in een tekst. Het totale aantal *tokens* in een tekst komt dus overeen met het totale aantal woorden in die tekst. Het aantal *types* in die tekst komt overeen met het aantal verschillende woorden.

Bij de selectie van de scholen werd rekening gehouden met de geografische spreiding, het onderwijsnet waartoe ze behoren en met het aantal allochtone leerlingen dat de school bezoekt (zie tabel 1).

% allochtonen		lokatie		net	
< 5%	2	Antwerpen	4	Stedelijk	5
5-15%	4	Brabant	6	Vrij	9
15-30%	3	Limburg	6	Gemeensch.	4
30-60%	5	Oost-Vlaanderen	2		
>60%	4				
totaal	18 scholen				

Tabel 1 Onderzoeksscholen basisonderwijs

Naast dit mondelinge corpus werd ook een corpus van 52 op de onderzoeksscholen gebruikte schoolboeken aangelegd. In dit corpus van 691.788 *tokens* werden zowel taalmethodes als reken- en zaakvakmethodes (aardrijkskunde, geschiedenis, natuurfkennis, verkeer, wereldoriëntatie en godsdienst) opgenomen.

### 2.2. Samenstelling van de woordenlijst

De woordenlijst die we in dit onderzoek hanteren, is een ijkinstrument voor de toegankelijkheid van het taalaanbod op school. De woorden die in de lijst zijn opgenomen, zijn geselecteerd op frequentie. Bij het gebruik van woordenlijsten samengesteld op basis van frequentiegegevens, moet rekening gehouden worden met



de betrouwbaarheid en representativiteit van de lijst. De betrouwbaarheid van een lijst staat in direct verband met de omvang van het corpus waarvan uitgegaan wordt. Hoe groter het corpus, hoe kleiner de kans dat een woord toevallig in de lijst terechtkomt. Doorgaans wordt een frequentie van 35 (Frumkina, 1963) of 40 (Alekseev, 1984) gehanteerd als betrouwbaarheidsgrens.

Binnen de beschikbare tijd konden we een corpus van 1.377.823 *tokens* verzamelen. Dit betekent dat we met een minimumfrequentie van 40 aan een woordenlijst van ongeveer 2.800 lemma's zouden komen. Het is zeer de vraag of een lijst van deze omvang uitgebreid genoeg is om als beschrijving van vakoverstijgende schooltaal te fungeren. Maar hoe groot moet de woordenlijst dan wel zijn?

Een hypothetische taalgebruiker die alle woorden uit de lijst kent, zou zonder al te veel moeilijkheden het taalaanbod op de basisschool moeten kunnen verwerken. Om de verwerkbaarheid van taalaanbod in te schatten maken we in dit onderzoek gebruik van dekkingspercentages. Een tekst die alle woorden bevat die in een bepaalde lijst staan, wordt volledig door die lijst gedekt en heeft een dekkingspercentage van 100%. Wanneer 25% van de woorden uit die tekst niet in de lijst voorkomen dan is het dekkingspercentage voor die tekst uiteraard 75%. Dekkingspercentages geven een vrij ruw beeld van de lexicale complexiteit van een tekst. Uiteraard bepalen ook andere factoren dan het aantal laagfrequente (of onbekende) woorden mee de (on)toegankelijkheid van een tekst. Dekkingspercentages hebben echter het voordeel dat ze op een snelle en doorzichtige manier informatie geven over de moeilijkheidsgraad van een tekst. De relatie tussen tekstdekking en tekstbegrip is niet eenduidig. Wanneer we een woordenlijst gelijk stellen aan het lexicon van een taalgebruiker dan geven de dekkingspercentages een beeld van het aantal voor die taalgebruiker bekende woorden uit een bepaalde tekst. Op die manier geven ze ook een aanwijzing omtrent het begrip dat die taalgebruiker aan een tekst kan toekennen. Wanneer iemand 80% van de woorden uit een tekst begrijpt, betekent dit echter in geen geval dat hij ook 80% van die tekst begrijpt. Over de relatie tussen tekstdekking en tekstbegrip of de ideale verhouding tussen bekende en onbekende (laagfrequente) woorden in een tekst is niet iedereen het eens. Sciarone (1979) stelt dat pas bij een dekking van 80% een begin van tekstbegrip optreedt terwijl pas bij 90% sprake is van een 'redelijk goed begrip'. Schouten-van Parreren (1979) haalt Klyc'nikova aan die voor het 'globaal begrip' van een literaire tekst een dekking van 75% vereist acht. Bij 90% bekende woorden zou men alle hoofdzaken begrijpen en bij 95% de meeste details. Nienhuis (1991) beschrijft twee experimenten waarbij tekstbegripsvragen gesteld werden over Nederlandse, Franse en Engelse teksten waaruit 10% en 25% van de meest frequente woordtokens verwijderd waren. Uit beide experimenten blijkt dat toch enig voorbehoud gemaakt moet worden bij de stelling dat bij een dekking van 90% 'alle hoofdzaken' begrepen worden. Zelfs van een 'redelijk goed begrip' blijkt nauwelijks sprake te zijn. Bij teksten met een dekkingspercentage van 75% was er in deze experimenten nog nauwelijks sprake van globaal begrip. Deze observaties geven steun aan de conclusies van Kameenui *et al.* (1982) die besluiten dat het voorkomen van één onbekend woord op veertien (wat overeenkomt met een dekkingspercentage van ongeveer 93%) al tot problemen leidt bij het beantwoorden van tekstbegripvragen die meer dan een letterlijke reproductie van de tekst vereisen. Dit alles in aanmerking

genomen, lijkt het wenselijk dat onze hypothetische taalgebruiker op zijn minst 90% van de woorden uit het taalaanbod waarmee hij geconfronteerd wordt begrijpt. Met andere woorden: onze uiteindelijke woordenlijst zou ongeveer 90% van het verzamelde taalaanbod moeten dekken.

Hazenberg (1993) geeft een overzicht van verschillende in de literatuur genoemde cijfers over de relatie tussen frequentie en tekstdekking. De cijfers die verschillende auteurs geven, blijken sterk uiteen te lopen. Voor het Nederlands wordt voor een dekking van 90% bij Deville (1985) een woordenschat bestaande uit de 3.500 meest frequente items verondersteld, terwijl Ostyn & Godin (1985) om tot dezelfde dekking te komen al 5.000 woorden nodig hebben. Hazenberg & Hulstijn (1992) onderzochten de dekking van een lijst van 23.550 *tokens* ten opzichte van het INL-corpus dat uit ruim 42 miljoen *tokens* bestaat. Uit hun gegevens blijkt dat de eerste 2.000 woorden (die in het INL-corpus een frequentie hoger dan 1.500 hebben) nauwelijks 80% van dit corpus dekken. Zij constateren ook dat er grote tekstspecifieke verschillen zijn. Met dezelfde 1.600 meest frequente woorden wordt een prozatekst voor 85,1% gedekt, een krantartikel voor 79,4% en een studietekst voor 88,5%. Eens te meer blijkt het belang van de representativiteit van een corpus: krantenartikels blijken in het INL-corpus ondervertegenwoordigd. Ook Nieuwborg (1992) toont aan dat de dekking die met een bepaalde lijst gehaald wordt sterk afhankelijk is van de tekstsoort. Dit is meteen een verklaring voor de sterk fluctuerende cijfers bij de verschillende auteurs. Hazenberg (1993) wijst op andere mogelijkheden tot 'manipulatie' van tekstdekkingscijfers: eigennamen en doorzichtige samenstellingen kunnen buiten beschouwing gelaten worden, woorden kunnen in meer of mindere mate gelemmatiseerd worden. Op zich is tegen deze operaties niet veel in te brengen, het probleem is enkel dat vele auteurs geen melding maken van de operaties die hun corpora ondergaan hebben, zodat niet duidelijk wordt waar hun dekkingspercentages precies op slaan en een onderlinge vergelijking moeilijk wordt. In dit verband is vooral belangrijk of er met *type*- of *tokendekking* gewerkt wordt. Onbekende woorden zijn vaak laagfrequent waardoor bij telling van het aantal *tokens* het aantal onbekende woorden kleiner is dan bij telling van het aantal *types* en de *tokendekking* bijgevolg hoger ligt. Overigens gaan de meeste auteurs uit van een *tokendekking*.

Hoe dan ook, met handhaving van een betrouwbaarheidsgrens van 40 en een lijst van 2.800 woorden komen we waarschijnlijk niet tot een woordenlijst die groot genoeg is om een voldoende tekstdekking - minimaal 90% - voor elk corpusonderdeel te waarborgen. Er werd dan ook afgestapt van een strenge betrouwbaarheidsgrens. Het frequentie criterium voor opname in de lijst werd versoepeld tot 10, wat zou leiden tot een lijst van ongeveer 5.000 woorden - in het licht van de hierboven genoemde cijfers een veilige marge. Het is duidelijk dat we hiermee toegevingen hebben gedaan op het vlak van de betrouwbaarheid van de woordenlijst als zodanig. Het gaat ons echter niet in de eerste plaats om het opstellen van een woordenlijst die rechtstreeks vertaalbaar is in doelstellingen voor het woordenschatonderwijs. We hebben enkel een referentiepunt nodig dat een beeld kan geven van de lexicale complexiteit van schools taalaanbod.

Alle woordvormen die een frequentie van minstens tien hadden, werden dus zonder meer als lemma in de lijst opgenomen. Meervoudsvormen, verkleinwoorden

en verschillende werkwoordsvormen werden herleid tot hun grondwoord, behalve als ze een afwijkende betekenis hadden (*brood* naast *broodje* bijvoorbeeld). De woordvormen met een frequentie tussen 5 en 10 werden handmatig gelemmatiseerd en als de frequentie van alle voorkomende vormen samen (de frequentie van het *type*) boven de 10 lag, werd het lemma in de lijst opgenomen. De woordenlijst moet een beeld geven van de taal waarin leerinhouden verpakt zijn en is dus vakoverstijgend. Woorden die behoren tot de specifieke woordenschat van één bepaald vak zijn voor onze doeleinden niet interessant. Om te vermijden dat een aantal vakspecifieke woorden toch toevallig in de lijsten terecht zou komen, werd een eenvoudig spreidingscriterium gehanteerd: woorden die slechts in één vak of in één leerjaar voorkwamen werden niet opgenomen. Eigennamen en namen van dagen, maanden en getallen werden ook niet opgenomen. Op basis van deze selectiecriteria werd de totale frequentielijst gereduceerd tot een lijst van 5.182 lemma's.

De lijst van alle geselecteerde woorden werd vervolgens in zes deellijsten opgesplitst. Elke deellijst komt overeen met één van de zes leerjaren in het Vlaamse basisonderwijs. Het eerste leerjaar in Vlaanderen komt overeen met groep 3 in Nederland. In principe werd een woord opgenomen in de lijst voor het leerjaar waarin het voor het eerst voorkwam. Om voor opname in een bepaalde deellijst in aanmerking te komen, moest een woord in dat leerjaar wel minstens drie keer voorkomen. Hierop werd een uitzondering gemaakt als een woord in twee verschillende vormen voorkwam. Als een woord in een bepaald leerjaar slechts twee maal in één vorm voorkwam, werd het enkel aan dat leerjaar toegewezen als het zowel in het mondelinge als in het schriftelijke corpus voorkwam. In tabel 2 is de verdeling van de lemma's over de verschillende deellijsten weergegeven.

deellijst voor leerjaar	1	2	3	4	5	6
aantal nieuwe lemma's	1583	782	1074	545	661	537
cumulatief aantal lemma's	1583	2365	3439	3984	4645	5182

Tabel 2 Verdeling van lemma's over basiswoordenlijst

De basiswoordenlijst, bevat voor een bepaald leerjaar de nieuwe lemma's voor dat leerjaar plus de lemma's uit de voorgaande leerjaren. Het aantal lemma's dat voor selectie in aanmerking komt, neemt sterk af in de loop van het basisonderwijs. Deze afname gebeurt niet echt geleidelijk. Tussen het derde en het vierde leerjaar ligt een duidelijke breuk. Het aantal lemma's dat in het vierde leerjaar aan de lijst toegevoegd kan worden, is ongeveer gehalveerd ten opzichte van het derde leerjaar.

De afname van het aantal nieuwe lemma's is typisch voor een woordenlijst die op frequentiegegevens gebaseerd is. Voor elk corpus geldt immers dat een relatief klein aantal woorden erg hoge frequenties haalt, terwijl een erg groot aantal woorden heel lage frequenties haalt. Dit verschijnsel staat bekend als de 'Wet van Zipf'. Een typische frequentielijst bestaat voor ongeveer 40% uit woorden die maar één keer voorkomen. Onder de woorden die in de latere leerjaren nieuw zijn zitten dan ook

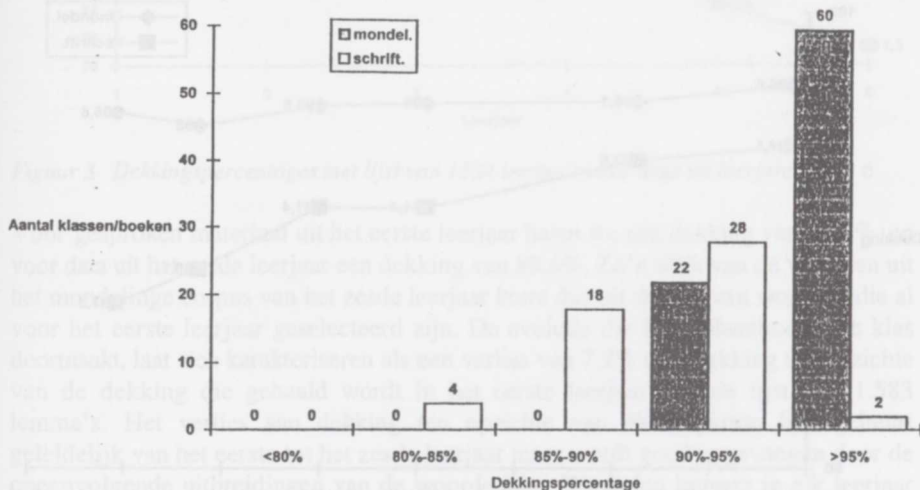


vooral woorden met een frequentie onder de 10 die niet voor selectie in aanmerking komen.

### 2.3. Tekstdekking voor corpusonderdelen

Voor elk corpusonderdeel - dat wil zeggen voor elk boek en voor elke mondelinge les - werd het dekkingspercentage berekend op basis van de voor dat leerjaar geselecteerde woorden. Voor teksten uit het derde leerjaar werden dus de deellijsten tot en met de lijst van het derde leerjaar gebruikt. Dit gebeurde als volgt. Een tekst werd door de computer vergeleken met alle voorkomende woordvormen in de woordenlijst. Tokens die niet in de lijst voorkwamen, werden beoordeeld door de onderzoekers. Persoonsnamen en getallen of ontbrekende vormen van lemma's die wel in de lijst stonden werden bij de 'bekende' woorden geteld. De resterende woorden werden als 'onbekend' beschouwd.

In figuur 1 geven we een overzicht van de dekkingspercentages die we op deze manier haalden.



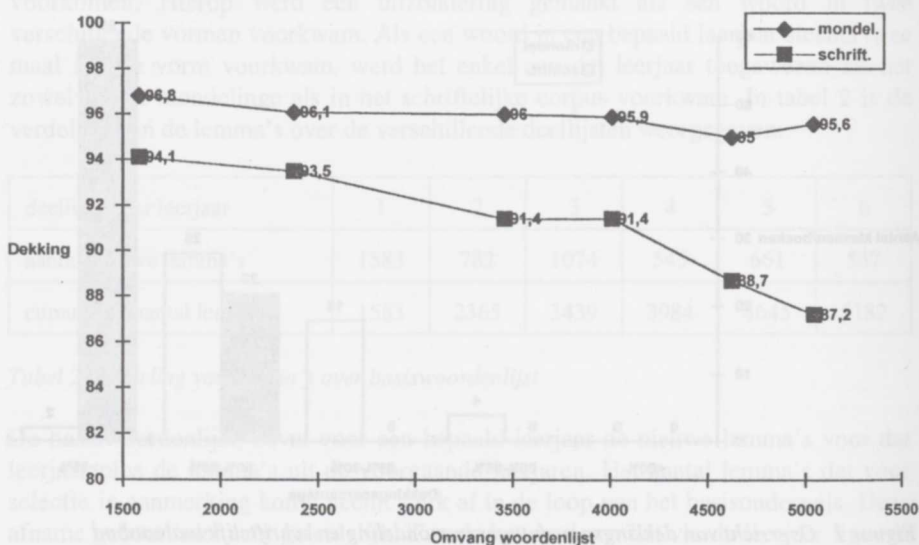
Figuur 1 Overzicht van dekkingspercentages in mondeling en schriftelijk taalaanbod

Het grote verschil tussen het mondelinge en het schriftelijke taalaanbod springt onmiddellijk in het oog. De dekkingspercentages liggen bij het mondelinge materiaal gemiddeld veel hoger. Bovendien is er tussen de verschillende klassen veel minder variatie dan tussen de verschillende boeken. In 22 van de 82 klassen wordt een dekking tussen de 90 en 95% bereikt. In maar liefst zestig klassen wordt zelfs een dekking van meer dan 95% bereikt. Een leerling die alle woorden kent die in een bepaald leerjaar in de lijst staan, zou weinig problemen mogen hebben om het

mondelinge taalgebruik in de klas te volgen. Daar staat tegenover dat vier schoolboeken niet de 85% halen die volgens Sciarone (1979) nodig is om tot een globaal tekstbegrip te komen. 18 boeken halen tussen de 85% en de 90% - niet genoeg om in Sciarone's termen van 'redelijk goed begrip' te kunnen spreken.

Het grote verschil in dekkingpercentages tussen het mondelinge en het schriftelijke taalaanbod is voor een deel te wijten aan de aanwezigheid van een groot aantal vakspecifieke woorden in schoolboeken. Dit soort woorden is bewust uit de lijsten geweerd. De laagste dekkingpercentages worden gehaald door teksten die veel van dit soort woorden bevatten: een aardrijkskundemethode voor het vijfde en zesde leerjaar en een geschiedenisboek voor het zesde leerjaar. Het vierde boek dat geen 85% haalt, is opvallend genoeg een taalmethode voor het vijfde leerjaar. De twee boeken die meer dan 95% halen zijn methodes voor aanvankelijk lezen in het eerste leerjaar.

In figuur 2 relateren we het gemiddelde dekkingpercentage per leerjaar aan het aantal lemma's in de woordenlijst. Elk punt op de grafiek komt overeen met het dekkingpercentage dat in de leerjaren 1 (met 1.584 lemma's) tot en met 6 (met 5.048 lemma's) gehaald wordt.



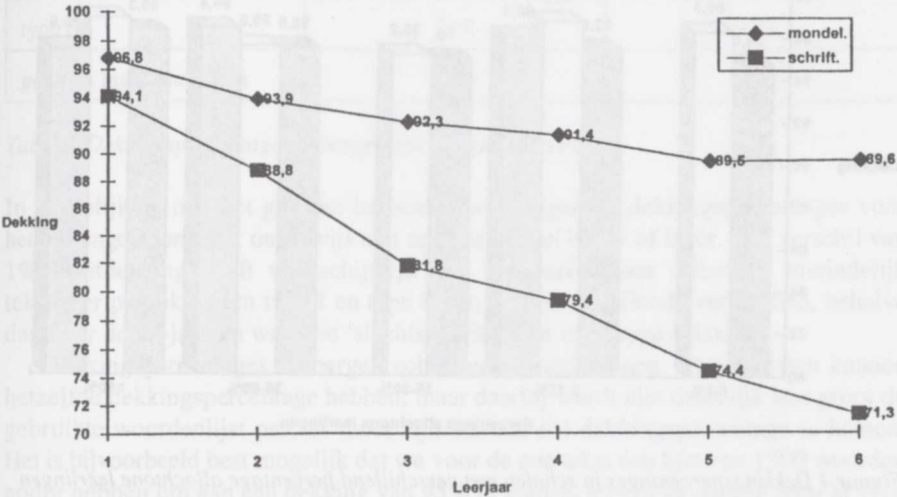
Figuur 2 Evolutie dekkingpercentages van leerjaar 1 tot 6 met lijst van toenemende omvang

Voor het mondelinge taalaanbod halen we maximaal 96,8% in het eerste leerjaar en minimaal 95% in het vijfde leerjaar. De dekkingpercentages schommelen nauwelijks. De dekkingpercentages voor het schriftelijke taalaanbod nemen sterk af naar het einde van het basisonderwijs toe. In het tweede leerjaar halen we nog 93,5%, in het zesde leerjaar komen we nog maar aan 87,2%. Uit dit evolutiepatroon kunnen we



afleiden dat kennis van de woorden uit de lijsten vanaf pakweg het vierde leerjaar niet voldoende is om in die leerjaren tot een voldoende begrip van schoolboekteksten te komen.

Het mondelinge taalaanbod in het eerste leerjaar wijkt veel minder sterk af van het aanbod in het zesde leerjaar dan we op het eerste gezicht zouden verwachten. Dat blijkt uit figuur 3, waarin we de dekkingspercentages opgenomen hebben die we in elk leerjaar halen met de lijst van 1.583 lemma's (de lijst voor het eerste leerjaar).

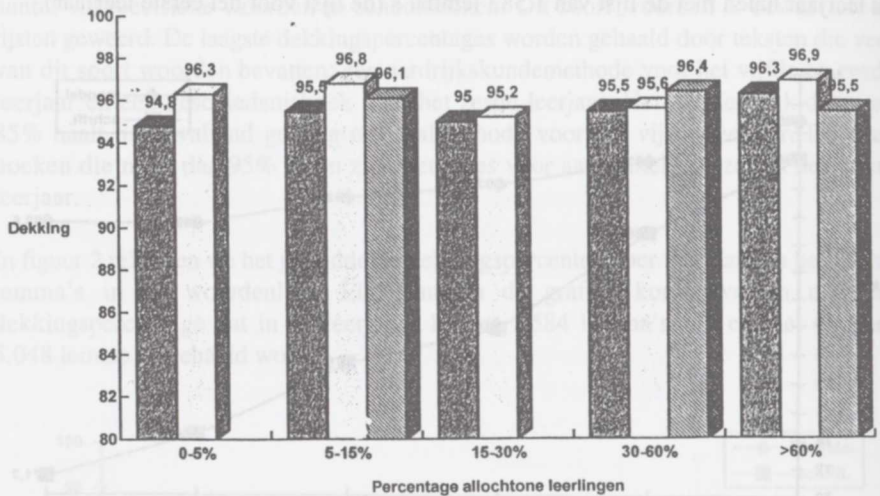


Figuur 3 Dekkingspercentages met lijst van 1584 lemma's voor data uit leerjaren 1 tot 6

Voor gesproken materiaal uit het eerste leerjaar halen we een dekking van 96,8%, en voor data uit het zesde leerjaar een dekking van 89,6%. Zo'n 90% van de woorden uit het mondelinge corpus van het zesde leerjaar komt dus uit de lijst van woorden die al voor het eerste leerjaar geselecteerd zijn. De evolutie die het taalaanbod in de klas doormaakt, laat zich karakteriseren als een verlies van 7,2% tekstdekking ten opzichte van de dekking die gehaald wordt in het eerste leerjaar met de lijst van 1.583 lemma's. Het verlies aan dekking ten opzichte van die beperkte lijst gebeurt geleidelijk van het eerste tot het zesde leerjaar maar wordt goed opgevangen door de opeenvolgende uitbreidingen van de woordenlijst. We halen immers in elk leerjaar minstens 95% met de lijst die voor dat leerjaar bedoeld is (zie figuur 2).

Voor het schriftelijke materiaal liggen de dekkingspercentages met de beperkte lijst uiteraard een stuk lager. Het totale verlies aan dekking bedraagt hier 22,8%. Het grootste verlies (7%) situeert zich in het derde leerjaar. Vanaf het vierde leerjaar zakken de percentages onder de 80%. Blijkens de gegevens uit figuur 2 slagen we er voor dit materiaal minder goed in om het verlies aan dekking op te vangen door nieuwe woorden aan de lijst toe te voegen.

Om na te gaan of er een verband is tussen het aantal 'moeilijke' woorden dat een leerkracht gebruikt en het aantal allochtone leerlingen in de klas bekijken we de gemiddelde dekkingspercentages in 13 onderzoeksscholen. De scholen zijn per twee of drie gegroepeerd, volgens het percentage allochtone leerlingen dat de school bezoekt.



*Figuur 4 Dekkingspercentages in scholen met verschillend percentage allochtone leerlingen*

Het verschil tussen de school met het hoogste en het laagste gemiddelde dekkingspercentage bedraagt 2%. Het lagere gemiddelde dekkingspercentage voor deze laatste school (94,9%) wordt echter veroorzaakt door een afwijkend laag percentage in één klas. Van grote verschillen tussen de verschillende scholen is geen sprake, laat staan van verschillen gekoppeld aan het percentage allochtonen dat de school bezoekt.

### 3. Vergelijking met andere corpora

#### 3.1. Buitengewoon onderwijs

Leerkrachten die lesgeven in het buitengewoon onderwijs beweren vaak dat ze hun woordgebruik aanpassen aan de lage taalvaardigheid van hun leerlingen. We onderzochten dan ook het taalaanbod op een aantal scholen voor buitengewoon lager onderwijs. Meer in het bijzonder keken we naar het taalaanbod in klassen voor kinderen met leermoeilijkheden (type 8) en kinderen met een lichte mentale handicap (type 1). Voor elk van deze onderwijstypes verzamelden we materiaal in twee leeftijdsgroepen (9-jarigen en 12-jarigen) op 6 verschillende scholen. Er werden 85.240 woordtokens verzameld in het type 1 en 74.735 in het type 8.

Voor iedere groep berekenden we het dekkingspercentage op basis van de lijsten voor respectievelijk het derde en het zesde leerjaar van het basisonderwijs. De resultaten zijn samengevat in tabel 3 met ter vergelijking de gemiddelde dekkingspercentages voor het derde en het zesde leerjaar uit het gewone basisonderwijs.

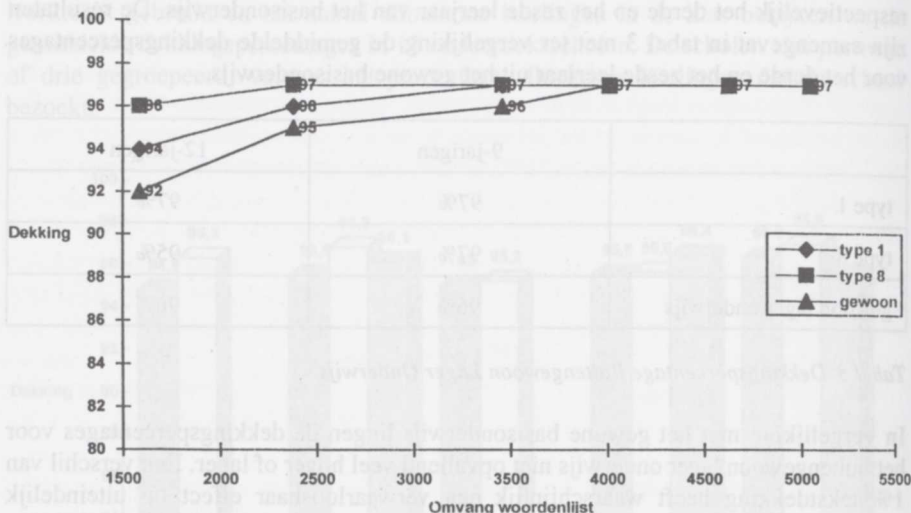
	9-jarigen	12-jarigen
type 1	97%	97%
type 8	97%	95%
gewoon lager onderwijs	96%	96%

Tabel 3 Dekkingspercentage Buitengewoon Lager Onderwijs

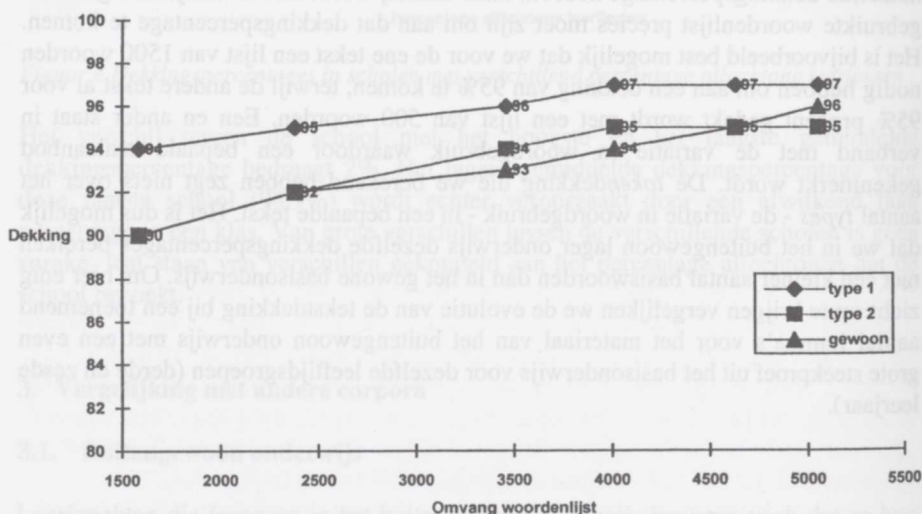
In vergelijking met het gewone basisonderwijs liggen de dekkingspercentages voor het buitengewoon lager onderwijs niet opvallend veel hoger of lager. Een verschil van 1% tekstdekking heeft waarschijnlijk een verwaarloosbaar effect op uiteindelijk tekstbegrip. Ook tussen type 1 en type 8 zijn er geen opvallende verschillen, behalve dan voor de 12-jarigen waar we 'slechts' 95% halen in de type 8-klassen.

Dekkingspercentages verbergen echter een aantal dingen. Twee teksten kunnen hetzelfde dekkingspercentage hebben, maar daarbij wordt niet duidelijk hoe groot de gebruikte woordenlijst precies moet zijn om aan dat dekkingspercentage te komen. Het is bijvoorbeeld best mogelijk dat we voor de ene tekst een lijst van 1500 woorden nodig hebben om aan een dekking van 95% te komen, terwijl de andere tekst al voor 95% procent gedekt wordt met een lijst van 500 woorden. Een en ander staat in verband met de variatie in woordgebruik waardoor een bepaald taalaanbod gekenmerkt wordt. De *tokendekking* die we berekend hebben zegt niets over het aantal *types* - de variatie in woordgebruik - in een bepaalde tekst. Het is dus mogelijk dat we in het buitengewoon lager onderwijs dezelfde dekkingspercentages bereiken met een kleiner aantal basiswoorden dan in het gewone basisonderwijs. Om hier enig zicht op te krijgen vergelijken we de evolutie van de tekstdekking bij een toenemend aantal lemma's voor het materiaal van het buitengewoon onderwijs met een even grote steekproef uit het basisonderwijs voor dezelfde leeftijdsgroepen (derde en zesde leerjaar).





Figuur 5a Dekkingspercentages mondelinge data 12-jarigen/6e leerjaar met lijsten van verschillende omvang



Figuur 5b Dekkingspercentages mondelinge data 9-jarigen/3e leerjaar met lijsten van verschillende omvang

Met de 1.584 woorden die zijn opgenomen in de lijst voor het eerste leerjaar van het basisonderwijs, bereiken we voor het materiaal uit het derde leerjaar van het gewone basisonderwijs een dekking van 92%. Met datzelfde aantal woorden komen we voor het materiaal uit de type 1-klassen aan 94% en voor de type 8-klassen aan 96%. Wat

het type 1 betreft, zet deze tendens zich ook door in het materiaal voor de 12-jarige kinderen. Waar we in het gewone basisonderwijs met 1.584 woorden aan 90% komen, halen we voor het type 1 materiaal al 94%. De type 8-klassen wijken voor deze groep dan weer nauwelijks af van het gewone basisonderwijs. Met uitzondering van deze laatste groep, kunnen we stellen dat we voor het buitengewoon onderwijs met minder woorden aan hogere dekkingpercentages komen.

Daar staat tegenover dat er doorgaans even veel woorden gebruikt worden die niet in de lijst staan. Voor de leerjaren/groepen waar het hier om gaat (3e leerjaar/9-jarigen en 6e leerjaar/12-jarigen) bevat de woordenlijst respectievelijk 3.458 en 5.048 lemma's. De dekkingpercentages die met lijsten van die omvang gehaald worden liggen - in de leeftijdsgroepen waarvoor de lijsten bedoeld zijn - zeer dicht bij elkaar. Een leerling uit de groep 9-jarigen (type 1 en 8) hoort in de klas dus ongeveer evenveel laagfrequente woorden als een leerling uit het derde leerjaar van het gewone basisonderwijs: 3 à 4% van het taalaanbod bestaat uit woorden die niet in de lijst zijn opgenomen. Voor de 12-jarigen schommelt het percentage laagfrequente woorden rond de 4%.

Om te achterhalen wat er precies gebeurt in de klassen zouden de data diepgaander geanalyseerd moeten worden. Een mogelijkheid is dat leerkrachten uit het buitengewoon onderwijs de neiging hebben om erg veel te herhalen, wat zich uit in minder variatie in woordgebruik. Als deze herhalingsstrategie bedoeld is om het taalaanbod toegankelijker te maken, dan schiet ze haar doel misschien wel voorbij. Wat betreft het voorkomen van laagfrequente woorden in de taal van de leerkracht, verandert er immers niets.

### 3.2. Vergelijking met Nederlandse data

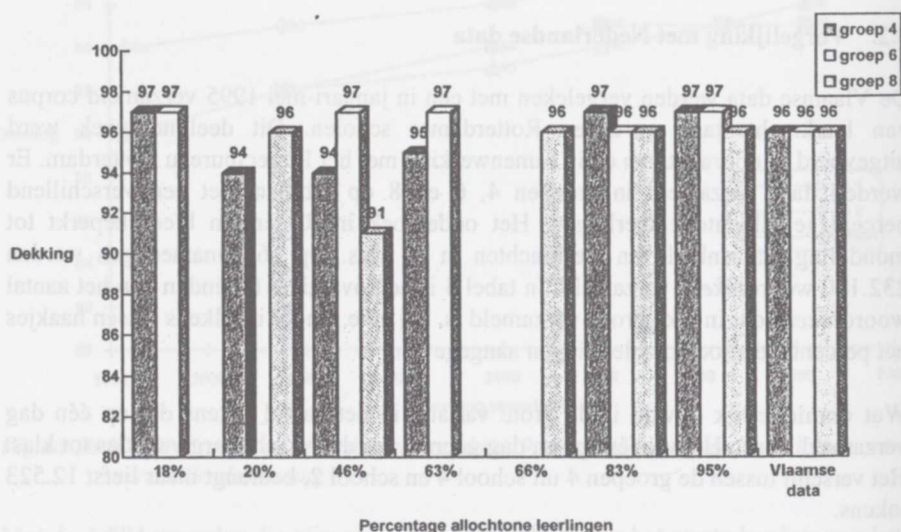
De Vlaamse data werden vergeleken met een in januari-mei 1995 verzameld corpus van leerkrachtentaal op zeven Rotterdamse scholen. Dit deelonderzoek werd uitgevoerd in opdracht van en in samenwerking met het Projectbureau Rotterdam. Er werden data verzameld in groepen 4, 6 en 8 op scholen met een verschillend percentage allochtone leerlingen. Het onderzoek in Rotterdam bleef beperkt tot mondeling taalaanbod van leerkrachten in de klas. Op 16 opnamedagen werden 232.140 woordtokens verzameld. In tabel 4 is een overzicht te vinden van het aantal woordtokens dat in elke groep verzameld is. Bij elke school is telkens tussen haakjes het percentage allochtone leerlingen aangegeven.

Wat onmiddellijk opvalt, is de grote variatie in het aantal tokens dat op één dag verzameld wordt. Hoeveel er op een dag gepraat wordt, verschilt erg van klas tot klas. Het verschil tussen de groepen 4 uit school 4 en school 2, bedraagt maar liefst 12.523 tokens.

	groep 4	groep 6	groep 8	totaal
school 1 (18%)	11813	13581	/	25394
school 2 (20%)	20451	/	10805	31256
school 3 (46%)	13715	17085	19580	50380
school 4 (63%)	7928	15277	/	23205
school 5 (66%)	/	/	16787	16787
school 6 (83%)	19769	18588	14369	52726
school 7 (95%)	8542	15013	8837	32392
totaal	82218	79544	70378	232140

Tabel 4 Aantal tokens per klas

Met de Vlaamse woordenlijst berekenen we de dekkingpercentages voor het Rotterdamse materiaal. De resultaten staan in figuur 6. Ter vergelijking hebben we ook het gemiddelde dekkingpercentage voor de overeenkomstige Vlaamse leerlingengroepen opgenomen.



Figuur 6 Dekkingpercentages Rotterdam per school/groep met percentage allochthone leerlingen

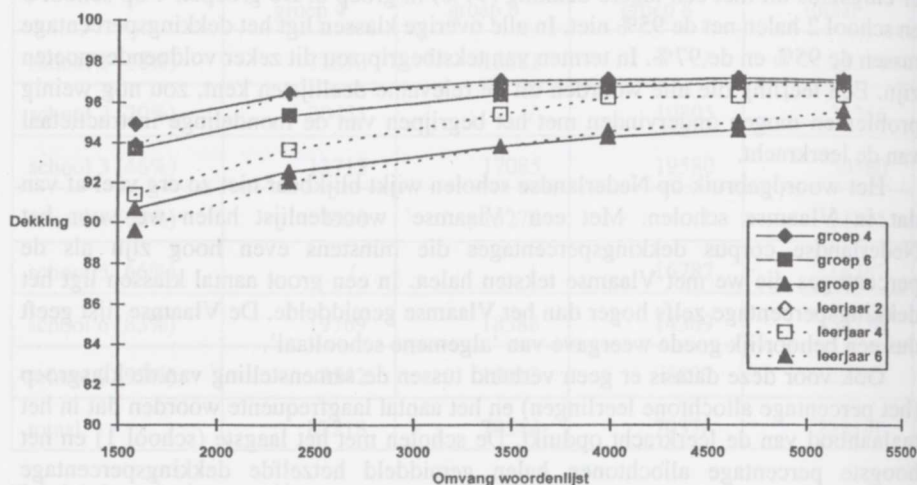


De Vlaamse woordenlijsten voor de verschillende leerjaren dekken een zeer groot deel van het taalaanbod in de Rotterdamse klassen. School 3 (de 46%-school) springt er enigszins uit met een lagere dekking (91%) in groep 8. De groepen 4 op school 3 en school 2 halen net de 95% niet. In alle overige klassen ligt het dekkingspercentage tussen de 95% en de 97%. In termen van tekstbegrip zou dit zeker voldoende moeten zijn. Een leerling die alle woorden uit de relevante deellijsten kent, zou nog weinig problemen mogen ondervinden met het begrijpen van de mondelinge instructietaal van de leerkracht.

Het woordgebruik op Nederlandse scholen wijkt blijkbaar niet zo erg veel af van dat in Vlaamse scholen. Met een 'Vlaamse' woordenlijst halen we voor het Nederlandse corpus dekkingspercentages die minstens even hoog zijn als de percentages die we met Vlaamse teksten halen. In een groot aantal klassen ligt het dekkingspercentage zelfs hoger dan het Vlaamse gemiddelde. De Vlaamse lijst geeft dus een behoorlijk goede weergave van 'algemene schooltaal'.

Ook voor deze data is er geen verband tussen de samenstelling van de klasgroep (het percentage allochtone leerlingen) en het aantal laagfrequente woorden dat in het taalaanbod van de leerkracht opduikt. De scholen met het laagste (school 1) en het hoogste percentage allochtonen halen gemiddeld hetzelfde dekkingspercentage (97%). Verder liggen de dekkingspercentages erg dicht bij elkaar. Dit bevestigt de resultaten uit het Vlaamse onderzoek. Ondanks het feit dat leerkrachten vaak beweren dat ze hun taalgebruik aanpassen aan de veronderstelde taalvaardigheid van de leerlingen is er van die aanpassingen op lexicaal niveau weinig terug te vinden. Leerkrachten gebruiken gewoon de woorden die ze nodig hebben om hun boodschap over te brengen.

In figuur 7 relateren we opnieuw het aantal lemma's in de woordenlijst aan de dekkingspercentages die we in elke groep bereiken. Uit deze gegevens kunnen we ook afleiden of de verdeling van de geselecteerde woorden over de lijsten voor de verschillende leerjaren/groepen ook voor de Rotterdamse scholen efficiënt is. Een goede verdeling van woorden over de verschillende deellijsten impliceert dat we bijvoorbeeld voor groep 6 een sterke stijging van het dekkingspercentage waarnemen tot de lijst voor het vierde leerjaar (545 lemma's) wordt toegevoegd. Op dat moment moet het dekkingspercentage hoog genoeg zijn. Daarna zou de tekstdekking niet meer in belangrijke mate mogen toenemen. Als dat toch gebeurt, betekent dit immers dat er ook in de volgende lijsten nog een groot aantal woorden staan die veel in groep 6 gebruikt worden.



Figuur 7 Evolutie dekkingspercentages in verschillende leerlinggroepen (Rotterdam en Vlaanderen) met lijsten van verschillende omvang

Met 1.583 lemma's halen we ook voor deze data in alle groepen al hoge dekkingspercentages. Het taalaanbod in groep 4 bestaat voor 94.9% uit de eerste 1.583 woorden uit de basiswoordenlijst, in groep 6 voor 93.7% en in groep 8 voor 90.7%. Het verschil tussen groep 4 en groep 6 is duidelijk kleiner dan het verschil tussen groep 6 en 8.

De verdeling van de lemma's over de lijsten (op basis van de criteria genoemd in paragraaf 2) is voor de groepen 4 en 8 in ieder geval efficiënt. Voor groep 4 zien we dat de dekking bij een verdere uitbreiding van de lijst tot 3.439 lemma's nog wel toeneemt met 0.7% maar hiervoor is wel een uitbreiding van de lijst met meer dan duizend lemma's vereist. Vanaf dan is de dekkingswinst verwaarloosbaar klein. Er komen geen nieuwe lemma's bij die voor het begrijpen van het taalaanbod in groep 4 van belang zijn. Voor groep 8 zien we dat het dekkingspercentage aanvankelijk relatief snel stijgt en vanaf de lijst voor het derde leerjaar (3.439) minder sterk toeneemt. Toch is de volledige lijst van 5.182 lemma's nodig om een dekkingspercentage van 95% te bereiken.

Een vergelijking met de Vlaamse resultaten leert ons dat wat groep 4 betreft de verschillen verwaarloosbaar klein zijn. In de Vlaamse tweede leerjaren halen we telkens ongeveer dezelfde dekkingspercentages met een lijst van dezelfde omvang. Ook het verschil tussen de groepen 8 en de Vlaamse zesde leerjaren is klein. In het algemeen is het zo dat het grootste deel van het taalaanbod, zowel in Vlaanderen als in Rotterdam, bestaat uit woorden die onder de eerste 1.583 lemma's van de basiswoordenlijst vallen. Voor beide corpora maken ze ongeveer 90% van het taalaanbod in groep 8/leerjaar 6 uit. De rest van het taalaanbod bestaat in Rotterdam

voor 4.5% uit woorden die onder de overige 3.599 lemma's vallen en verder voor 4.8% uit woorden die niet in de lijst staan. In Vlaanderen zijn die percentages respectievelijk 6.2% en 4.2%. Op dit punt liggen beide corpora dus erg dicht bij elkaar.

Wat groep 6 betreft, zijn de verschillen tussen Vlaanderen en Rotterdam iets groter. Voor het Vlaamse materiaal uit het vierde leerjaar halen we met 1.583 lemma's 91.4%. Met 1.583 lemma's komen we voor het Rotterdamse corpus veel verder. Materiaal uit groep 6 wordt voor 93.7% gedekt door de lijst van 1.583 lemma's. Een dekkingspercentage van 95% wordt in Rotterdam al bereikt met de lijst van 2.365 lemma's die voor het Vlaamse tweede leerjaar geselecteerd zijn. In Vlaanderen zijn hiervoor 3.458 lemma's nodig.

Aan deze observatie knopen we de volgende bedenkingen vast. In de loop van het basisonderwijs vindt zowel in Vlaanderen als in Rotterdam een verschuiving plaats wat betreft de moeilijkheidsgraad van het taalaanbod. Deze verschuiving behelst zowel in Vlaanderen als in Rotterdam een verlies van ongeveer 4% tekstdekking in het zesde leerjaar/groep 8 ten opzichte van de dekking die gehaald wordt met een lijst van 1.583 lemma's in het tweede leerjaar/groep 4. (Ten opzichte van data uit een Vlaams eerste leerjaar is dit een verlies van zo'n 7%.) Dit verlies wordt in beide gevallen goed opgevangen door de lemma's die aan de lijst worden toegevoegd. Als leerlingen de woorden uit de lijst voor een bepaald leerjaar op het moment dat ze in dat leerjaar zitten ook effectief kennen, dan bevat het taalaanbod dat ze in de klas te horen krijgen nog zo'n 4% onbekende woorden.

Het opvallendste verschil tussen Vlaanderen en Rotterdam zit hem echter in het moment waarop dit verlies aan dekking zich manifesteert. In Vlaanderen zien we het al opduiken in het vierde leerjaar. In het corpus met taalaanbod uit Vlaamse vierde leerjaren halen we slechts 91.4% met 1.583 lemma's. In Rotterdam halen we voor groep 6 met even veel lemma's nog 93.7%. Het verschil tussen groep 4 en groep 6 (1.2%) is maar half zo groot als het verschil tussen het tweede en het vierde leerjaar (2.5%). De evolutie naar 'moeilijker' taalaanbod, met meer variatie in woordgebruik, zet zich in Rotterdam blijkbaar iets later door, namelijk tussen groep 6 en groep 8. Op dat moment wordt de situatie omgekeerd. Het verschil tussen het vierde en het zesde leerjaar (1.8%) is bijna de helft kleiner dan het verschil tussen groep 6 en groep 8 (3%).

Van het Rotterdamse corpus werd een frequentielijst gemaakt die vervolgens werd vergeleken met de Vlaamse basiswoordenlijst. Eerst werd nagegaan welke woorden uit het Rotterdamse corpus met een frequentie hoger dan tien - de selectiegrens die bij het eerste onderzoek gehanteerd werd - niet in de Vlaamse lijst voorkwamen. De bereikte dekkingspercentages wijzen er echter al op dat deze lijst vrij beperkt zal zijn. Alles bij elkaar vonden we nauwelijks 19 woorden met een frequentie hoger dan tien die niet in de Vlaamse basiswoordenlijst voorkomen. Daarvan kwamen er 7 slechts in één enkele les voor.

Omdat het Rotterdamse corpus een stuk kleiner is hebben we ook gekeken naar ontbrekende woorden met een frequentie tussen 4 en 9. Deze woorden zouden bij uitbreiding van het corpus misschien een hogere frequentie kunnen halen. Deze versoepeling van het selectie criterium leidt tot een lijst van 106 woorden waarvan er



63 slechts in één klas voorkomen. Naast woorden die toevallig vaak genoeg voorkomen omdat ze met het les- of gespreksonderwerp te maken hebben (*digitaal, werkweek*), bevat deze lijst ook een aantal 'typisch Nederlandse' woorden die niet in de Vlaamse lijst staan (*hartstikke, gulden, schorten, prullenbak, grienen, zaklamp*). Voor een aantal van deze woorden zijn in Vlaanderen alternatieven meer gebruikelijk. Een *prullenbak* is een *vuilbak*, een *repetitie* is een *toets*, *schorten* is *schelen* en een *zaklamp* is *doorgaans*, dat wil zeggen buiten de context van een taalzuiveringsles, een *pillamp*. Alles bij elkaar zijn zo'n 15 woorden als 'typisch Nederlands' te kwalificeren.

De lijst van ontbrekende woorden bevat verder een klein aantal woorden die om een onduidelijke reden niet in de Vlaamse lijst staan maar die op intuïtieve basis toch als 'belangrijk om de instructietaal op school te begrijpen' gekwalificeerd kunnen worden. Het gaat om woorden als *afhalen, niveau, inmiddels* en *aanleiding*.

De lijsten werden ook in de omgekeerde richting vergeleken. We gingen na welke woorden uit de Vlaamse basiswoordenlijst niet in het Rotterdamse corpus voorkwamen. Gezien het verschil in omvang tussen beide corpora is de verwachting dat een relatief groot aantal lemma's zullen ontbreken in de Rotterdamse lijst. Van het totale aantal lemma's (5.182) komen er 2.619 (50.5%) niet in het Rotterdamse corpus voor. In die 50% zitten een aantal 'typisch Vlaamse' woorden, zoals *daarjuist, lessenaar, pennenzak, turnen, overschot*. Daarnaast ontbreken ook typische schrijftaalwoorden zoals *nimmer, sedert, ofschoon*. Het Rotterdamse corpus bevat immers geen schriftelijk materiaal. Daarom ontbreken ook een groot aantal algemene vaktaalwoorden, die wel frequent zijn in de schoolboeken maar door de beperktheid in de tijd van het onderzoek niet door de leerkracht in de klas gebruikt zijn. Bij uitbreiding van het corpus met schriftelijk materiaal zouden de verschillen allicht veel kleiner worden.

#### 4. Besluit

Het belangrijkste doel van dit onderzoek was het in kaart brengen van de lexicale taalvaardigheidseisen op de basisschool. De woordenlijsten die werden geproduceerd geven - samen met de gegevens over het gebruik van de lexicale items die ze bevatten - een beeld van die eisen. Het mondelinge taalaanbod op de basisschool kan worden gevat in een lijst van zo'n vijfduizend woorden die ongeveer 96% van dat taalaanbod dekken. Dit dekkingspercentage ligt een stuk hoger dan de omvang van de lijst zou doen vermoeden (zie paragraaf 2.2). Het karakter van de mondelinge schooltaal verandert niet echt. Een vrij beperkte basiswoordenschat die in de eerste twee leerjaren al aanwezig is, vormt het fundament. In het kader van een onderzoek naar de gebruikte taalfuncties in hetzelfde corpus constateerde Humblet (1993) al dat vooral taalfuncties die te maken hadden met de organisatie van het klasgebeuren en het regelen van de communicatie in de klas vaak voorkwamen. Die situatie en de woordenschat die gebruikt wordt om er naar te verwijzen verandert niet echt in de verschillende schooljaren. Dit is een mogelijke verklaring voor de weinig gevarieerde woordenschat in de klas.

Om het schriftelijke taalaanbod te begrijpen, is een veel grotere woordenschat nodig. Het probleem is dat die woordenschat zich erg moeilijk laat vatten in een woordenlijst. We kunnen op basis van frequentiegegevens niet genoeg woorden aan de lijst toevoegen om ervoor te zorgen dat de dekkingpercentages voor een groot deel van de schoolboeken hoog genoeg blijven. Dit betekent dat de woordenlijsten uit dit onderzoek (en misschien geldt dit wel voor alle woordenlijsten) slecht renderen als men ze zou beschouwen als een opsomming van leerdoelen.

Hiermee belanden we bij de vraag naar de bruikbaarheid van onze onderzoeksresultaten en woordenlijsten voor de onderwijspraktijk. Het verwerven van een voldoende grote woordenschat is geen doel op zich, maar staat in functie van het verwerven van de taalvaardigheid die nodig is om binnen een bepaald domein, i.c. het schoolse leren, te functioneren. Hoe het ook vorm wordt gegeven, goed woordenschatonderwijs is in de eerste plaats taalvaardigheidsonderwijs waarbij leerdoelen zorgvuldig geselecteerd zijn in overeenstemming met de behoeften van de doelgroep. Die leerdoelen kunnen op verschillende niveaus (lexicaal, taalfunctioneel) geformuleerd worden. De keuzes die op dit vlak gemaakt worden, leiden uiteindelijk naar verschillende vormen van - meer of minder expliciet - woordenschatonderwijs.

Expliciet woordenschatonderwijs vertrekt van een aantal doelstellingen op lexicaal niveau. Anders geformuleerd: er moet een aantal welomschreven woorden geleerd worden. Het is van groot belang dat de selectie van te leren woorden zeer zorgvuldig gebeurt, want het aantal woorden dat expliciet behandeld kan worden, is beperkt.

Vanuit dat perspectief bekeken stelt onze woordenlijst (en bij uitbreiding elke woordenlijst gebaseerd op frequentiegegevens) de expliciete methodeschrijver voor een aantal problemen. Aan de ene kant bevat ze - ondanks de wel erg soepele betrouwbaarheidscriteria - niet genoeg woorden om de lexicale behoeften van basisschoolleerlingen te dekken (vooral dan voor het schriftelijke taalaanbod in de hogere leerjaren). Aan de andere kant bevat ze een groot aantal woorden die voor een deel van de doelgroep waarschijnlijk al gekend zijn. Dit wordt duidelijk als we de omvang van onze woordenlijst afzetten tegen een aantal schattingen van de omvang van de woordenschat van onze doelgroep. Dijkstra & Kempen (1984) stellen dat zesjarigen een passieve woordenschat van 6.000 tot 12.000 woorden hebben. Verhoeven & Vermeer (1993) concluderen dat autochtone Nederlandse kinderen uit een laag sociaal milieu op het eind van het basisonderwijs ongeveer 15.000 woorden (receptief) kennen. Dit cijfer ligt in de buurt van de 14.740 woorden die Sanders (1990) geeft als zijnde de omvang van de receptieve woordenschat van brugklasleerlingen. De woordenschat van allochtonen op het eind van het basisonderwijs ligt volgens Verhoeven en Vermeer rond de 10.000 lexicale items. Wanneer we deze cijfers afzetten tegen de omvang van onze woordenlijst - ongeveer 5.000 lemma's - dan is het erg waarschijnlijk dat leerlingen uit het basisonderwijs al een groot deel van de woorden uit deze lijst zullen kennen, zeker als het gaat om de zeer hoogfrequente woorden uit de lijsten voor het eerste en tweede leerjaar. Coenen & Vermeer (1988) komen tot een lijst van 4.332 woorden die een beeld geeft van de actieve woordenschat van allochtone kinderen op de basisschool. Kinderen uit groep 3/4 gebruiken 1.819 woorden (hapaxen werden niet meegeteld). Op basis van een

steekproef hebben we vastgesteld dat ongeveer 14% van de woorden uit onze woordenlijst voor het eerste leerjaar niet in de lijst van Coenen en Vermeer voor de groepen 1 tot 4 voorkomen. Het gaat dan onder andere om typisch Vlaamse woorden (*efkens*) en woorden die typisch zijn voor de taal van de leerkracht in de klas (*afschrijven*). Dit alles wijst er op dat een groot aantal woorden uit de basiswoordenlijst voor het eerste leerjaar en tweede leerjaar (2.365 lemma's) waarschijnlijk al gekend zullen zijn.

Het woordselectieprobleem is dus zeker niet opgelost met de productie van een woordenlijst. Het risico blijft bestaan dat er heel veel tijd geïnvesteerd wordt in woorden die leerlingen al kennen en dat belangrijke woorden niet behandeld worden. In de hogere leerjaren wordt de selectie van woorden als leerdoelen steeds moeilijker. Nieuwe woorden zijn laagfrequent en zo sterk gebonden aan de onderwerpen die in de lessen ter sprake komen dat hun kwantitatieve belang niet meer objectief hard te maken is. Toch is het duidelijk dat er nog heel wat woorden geleerd moeten worden om het taalaanbod in die hogere leerjaren te verwerken.

In Vlaanderen is er dan ook voor gekozen om niet te vertrekken vanuit een vorm van expliciet woordenschatonderwijs, maar om de taken die leerlingen moeten uitvoeren centraal te stellen. Die taken vormen als het ware de natuurlijke context van de woordenschat die aan bod komt. Op het waarom van de keuze voor taakgericht onderwijs gaan we hier niet dieper in (zie hiervoor bijvoorbeeld Linsen, 1994). Een voordeel van een taakgerichte aanpak is dat een groot aantal woorden in relevante contexten aan bod kan komen. Bovendien kan er soepel ingespeeld worden op behoeften van leerlingen in verband met explicitering van woordbetekenissen. Het belangrijkste is dat leerlingen onmiddellijk relevante dingen moeten doen met woorden wat voor de retentie op lange termijn wellicht de meest gunstige conditie is.

Binnen een taakgerichte aanpak kunnen woordenlijsten een dubbele rol spelen. Aan de hand van de woordenlijsten kan bepaald worden hoe een taak zich verhoudt tot het taalaanbod dat leerlingen op een doorsnee schooldag te verwerken krijgen. Op die manier krijgen we een beeld van de moeilijkheidsgraad van een taak. Die moeilijkheidsgraad is van cruciaal belang voor het leereffect van een taak. Bovendien zorgen we door het systematisch integreren van belangrijke schooltaalwoordenschat in de taken dat deze hun doel niet voorbijschieten. Woordenschat wordt op die manier het controlemiddel bij uitstek voor pedagogische taken. Het gaat binnen een dergelijke aanpak dus zeker niet om het incidenteel verwerven van woorden die 'toevallig' aan bod komen. In de hele discussie over expliciet versus impliciet woordenschatonderwijs wordt dit laatste overigens al te vaak gelijkgesteld met incidenteel leren van woorden.

Welke invalshoek men ook kiest, het is van het grootste belang om niet uit het oog te verliezen dat dit soort woordenlijst enkel en alleen een indicatie geeft van hoe een talig domein eruit ziet. Het is zeker geen hapklare lijst leerdoelen. De talige behoeften van basisschoolleerlingen zijn niet zomaar uit te drukken in een statische woordenlijst.



## Bibliografie

- Alekseev, P.M. (1984). *Statistische Lexikographie*. Bochum: Brockmeyer.
- Coenen, M. & A. Vermeer (1988). *Nederlandse Woordenschat Allochtone Kinderen*. Tilburg: Zwijssen.
- Derville, G. e.a. (1985). *Measuring the F.L. learner's lexical needs*, Paper presented at the 5th European LSP Symposium, Leuven.
- Dijkstra, T. & G. Kempen (1984). *Taal in uitvoering. Een inleiding tot de psycholinguïstiek*. Groningen: Wolters-Noordhoff.
- Frumkina, R.M. (1963). The application of statistical methods in linguistic research. In: O.S. Akhmanova, A. Mel'chuk, R.M. Frumkina, E.V. Paducheva, *Exact Methods in linguistic research*. Berkeley: University of California Press.
- Hazenbergh, S. & J. Hulstijn (1992). Woorden op zicht. Wordselectie ten behoeve van het NT2-onderwijs. *Levende-Talen*, 467, 2-7.
- Hazenbergh, S. (1993). Tekstdekking: Goochelen met cijfers? *Levende Talen*, 476, 10-15.
- Humblet, I. (1993). Schoolse Taalfuncties in het basisonderwijs. *Spiegel*, 11, 2, 31-55.
- Kameenui, E.J., D.W. Carnine & R. Freschi (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly*, 17, 3, 367-88.
- Linsen, B. (1994). Met vallen en opstaan: de introductie van taakgericht taalonderwijs. Een praktijkbeschrijving uit het basisonderwijs in Vlaanderen. In: S. Kroon & T. Vallen (red.). *Nederlands als tweede taal in het onderwijs. Praktijkbeschrijvingen uit Nederland en Vlaanderen*. Nederlandse Taalunie Voorzeten 46. Den Haag: Sdu Uitgeverij Koninginnegracht.
- Nienhuis, L.J.A. (1991). Het begrip van teksten met 10% tot 25% onbekende woorden. *Toegepaste Taalwetenschap in Artikelen*, 41, 57-66.
- Nieuwborg, E. (1992). Tekstdekking en tekstbegrip. Een experimenteel onderzoek. In: André Halbo (red.) *Evaluation and Language Teaching, Liber Amicorum Frans van Passel*. Bern, etc.: Peter Lang.
- Ostyn, P. & P. Godin (1985). RALEX: an alternative approach to language teaching. *The modern Language Journal*, 69, 346-355.
- Sanders, M. (1990). Met andere woorden. Een onderzoek naar de omvang van de receptieve woordenschat van anderstalige brugklasleerlingen. *Levende Talen*, 455, 424-427.
- Sciarone, A. (1979). *Woordjes leren in het vreemde-talenonderwijs*. Muiderberg: Coutinho.
- Schouten-van Parreren, M.C. & C. van Parreren (1979). De verwerving van een vreemdtalige woordenschat. *Levende Talen*, 341, 250-270.
- Verhoeven L. & A. Vermeer (1993). *Taaltoets Allochtone Kinderen Bovenbouw*. Tilburg: Zwijssen.

