

Bijlage 1

Tabel 3. Gemiddeld aantal antwoorden per scenario in Experiment 1 (sd's tussen haken).

Scenario	Conditie			
	Oud & Nieuw	Oud&Nieuw+	Alleen-nieuw	Alleen-oud
Oud	0.2 (0.5)	0.0 (0.0)	0.0 (0.0)	2.8 (1.2)
Nieuw	3.2 (1.0)	3.7 (1.0)	3.6 (1.2)	0.2 (0.4)
Anderszins	2.6 (1.0)	2.3 (1.0)	2.4 (1.2)	3.0 (1.2)

Tabel 4. Gemiddeld aantal antwoorden per scenario in Experiment 2 (sd's tussen haken).

Scenario	Conditie			
	Zwak-oud	Zwak-Nieuw	Sterk-Nieuw	Sterk-oud
Oud	1.55 (1.53)	0.95 (1.43)	1.50 (1.35)	1.56 (1.58)
Nieuw	1.91 (1.07)	2.20 (1.32)	1.21 (1.02)	1.63 (1.24)
Anderszins	2.55 (1.10)	2.85 (1.04)	3.29 (1.16)	2.81 (1.39)

Samenvatting

Aan de hand van een structuuranalyse op de verklarende tekst van een twaalfjarige basisschoolleerlinge wordt uiteengezet hoe tekstanalyse met PISA (de Procedures voor Incrementele Structuur Analyse) bij kan dragen aan de theorievorming over de conceptuele processen tijdens het schrijven. Eerst wordt de validiteit van de PISA-structuur onderbouwd door aan te tonen dat de tekststructuur correspondeert met de (ideosyncratische) interpunctieconventies van de schrijfter. Daarna wordt uitgelegd hoe uit de tekststructuur strategieën en bijbehorende procedures gereconstrueerd kunnen worden. Evidentie voor de validiteit van deze 'cognitieve plannen' wordt verkregen uit de verdeling binnen de tekst van spelfouten, taalfouten en zelfcorrecties. Tot slot wordt ingegaan op de generaliseerbaarheid van de resultaten en op de wenselijkheid om 'off-line' methoden te combineren met 'on-line' technieken zoals pauzemetingen.

1. Inleiding

In de jaren tachtig is herhaaldelijk betoogd dat tekstanalyse een zeer veelbelovend onderzoeksparadigma zou kunnen zijn voor schrijfprocesonderzoek (zie bijv. Bereiter & Scardamalia, 1983; Cooper & Matsuhashi, 1983). Maar tot op heden beperken de resultaten zich tot een aantal nogal impressionistische studies (o.a. Witte & Cherry, 1986). Het is onze bedoeling om de tekst-analytische benadering van het schrijfprocesonderzoek te revitaliseren (Van der Pool, 1995, deze bundel; Sanders & Van Wijk, 1994, 1996a,b; Schilperoord, 1996, deze bundel; Van Wijk, 1992, 1995; Van Wijk & Sanders, 1991). Aan de hand van een gevalstudie verduidelijken wij eerst onze methode van tekstanalyse en daarna de wijze waarop wij mentale processen uit de geanalyseerde tekst reconstrueren.

De tekst in Tabel 1 is een uitleg van Sinterklaas die werd geschreven door een meisje uit groep acht van de basisschool. Deze tekst is geselecteerd omdat het de langste was in een bestand van 62 teksten. De leerlingen schreven een uitleg over een technologisch onderwerp (*de telefoon*) of een cultureel onderwerp (*Sinterklaas*). De instructie luidde: "In een vreemd land zit je een Nederlands tijdschrift te lezen. Een inboorling kijkt over je schouder mee en begint allerlei vragen te stellen over een artikel met veel foto's van *[Topic]*. Hij weet helemaal niets van *[Topic]* en jij begint hem het uit te leggen. *[Topic]* is ..". De tekst is opgedeeld in segmenten: hoofdzinnen hebben een volgnummer gekregen, afhankelijke deelzinnen een letter achter het nummer van hun bijbehorende hoofdzin (zie voor segmentatierregels, Sanders & Van Wijk, 1996a). We verwijzen naar segmenten met hun code voorafgegaan door een *s* (bijv. *s7a* en *s23*). Codes voor interpunctie en pagina-opmaak zijn voor het segment gezet, die voor fouten en zelfcorrecties erbinnen. De procesgerichte tekstanalyse verloopt in twee stappen: eerst van tekst naar structuur, en dan van structuur naar proces. Deze werkwijze wordt op de voorbeeldtekst geïllustreerd (zie Sanders & Van Wijk, 1996b, voor eenzelfde analyse van een uitleg van de telefoon geschreven door een

zeventienjarige jongen). Voor de plausibiliteit van beide analysestappen presenteren wij onafhankelijke evidentie: voor de eerste stap aan de hand van de interpunctieconventies van de schrijfstijl; voor de tweede stap aan de hand van de verdeling van fouten en zelfcorrecties.

Tabel 1. Voorbeeldtekst opgesplitst in segmenten en gecodeerd voor (1) interpunctie en opmaak, (2) spel- en taalfouten en (3) zelf-correcties.

0	Sinterklaas
1	Sinterklaas is een man die per boot
0	(zo'n ding wat op water is)
1a	naar Holland komt met zwarte pieten,
1	een zwarte piet is net zo als jou
2	want hij is ook zwart
3	en hij heeft ook een roe,
4	dat zijn allemaal takjes
5	Ook heeft hij een zak waar alles in kan
@	bijv als kinderen
7a	(dat zijn kleine mensjes)
0	stout
7a	(dat is niet leuk)
0	zijn,
7a	dan kunnen zij erin
7	Maar nu zitten er cadeautjes in
@	Cadeautjes zijn: als ik bijv een banaan aan iemand geef.
@	Zwarte piet loopt met sinterklaas op de (huizen-< huizen,
@	(daar doen de mensen alles in)
,	en dan gooien ze wat in de schoorsteen
0	dat is een ding waar rook uit komt.
.@	De kinderen zeten (SP) hun schoen voor de schoorsteen
15	en zingen een liedje
16a	Dan als ze uit bed komen
0	(een bed daar slaap je in)
16b	dan zit er iets in hun schoen.
16	Sinterklaas heeft ook een boek
.@	(dat is waar jij nou in zit te kijken)
0	en daar staat in of je stout bent geweest
@	Op 5 december is Sinterklaas jarig
0	(5 december is een dag
0	en als je jarig bent
0	krijg je cadeautjes)
0	maar Sinterklaas geeft cadeautjes weg.
.@	[Ze] Dan [dees< doen de Zwarte pieten cadeautjes in een zak
24	en (SY2 zij) bonsd (SP) (SY1) dan op een deur
25	(een dees< deur is iets dat voor (SY2 een) huis staat)
0	en dan zetten ze (SY2 een) (cadeu< cadeautje [in een zak] [op< voor de deur
27	en als iedereen zijn cadeautjes heeft
@	Dan neemt die daar woont ze mee
28	dan zegen (SP) ze dank u wel Sinterklaasje
29a	
29	

@	30	Daar (SY2 na) gaat Sinterklaas weer met zijn boot (SY2 naar) Spanje
@	31	Spanje is een land heel ver weg
@	32	De kinderen zijn dat (SP) meestal erg blij
0	33	(bij< blij dat is erg leuk of aardig)
@	34	Oh ja Sinterklaas rijdt ook op een paard
35		dat is een groot dier
35a		waar je op kan zitten
@	36	zo dat was het.
37		Sinterklaas heeft een hoed op zijn hoofd een rode mantel en (SY3) een witte (y< jurk en een staf
@	38	En zwarte piet heeft (SP) een korte broek en (SY3) een (hoed< hoed met een veer en een soort trui

Noot. Interpunctie en opmaak: . ; : segment volgt op een punt resp. komma; @ : segment begint op een nieuwe regel (en in de regel ook met een hoofdletter); () : segment staat tussen haakjes. Spel- en taalfouten: {SP} : spelfout; {SY1} : congruentiefout; {SY2 **} : element ** ontbreekt; {SY3} : nevenschikkend voegwoord werd te vroeg geplaatst. Zelf-correcties: (**< : correctie van **, [**< : vervanging van **, [**]: deletie van **

2. Stap 1: van tekst naar structuur

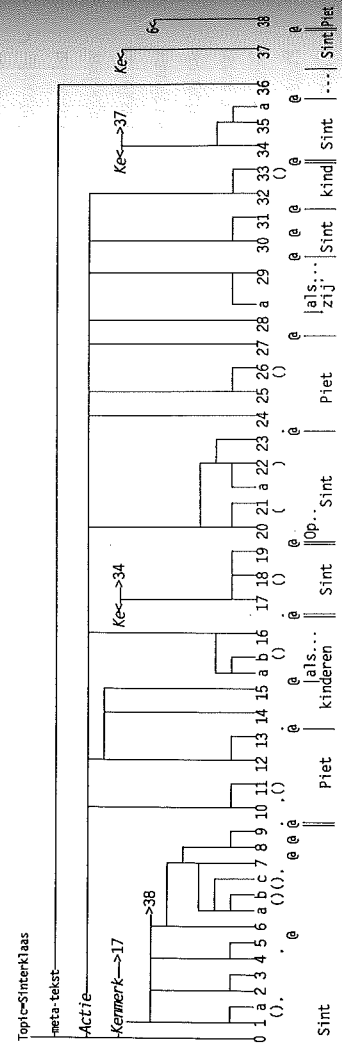
Ons onderzoek richt zich vooral op verklarende tekst, van eenvoudige beschrijvingen tot technische explicaties. Om zulke teksten objectief en betrouwbaar te analyseren hebben wij de Procedures voor Incrementele Structuur Analyse ontwikkeld, kortweg PISA (Sanders & Van Wijk, 1994, 1996a). De techniek bestaat uit een geordende verzameling van contextgevoelige *als-dan* regels (voor de meest recente 'update', zie Van der Pool, 1995: 193-207). PISA kent aan een tekst een hiërarchische en een relationele structuur toe in de vorm van een boomstructuur met een inhoudelijk label aan iedere verbinding. De analyse gaat uit van expliciet omschreven kennis van talige markerings, een paar tekstschema's en een kleine verzameling wereldkennis. De analyse verloopt incrementeel: segmenten worden achter elkaar verwerkt overeenkomstig hun volgorde in de tekst zonder een voorafgaande globale inspectie van de hele tekst. Een analyse begint altijd vanuit twee tekstschema's: de *Kenmerk-* en de *Handelingslijn*. De eerste lijn organiseert het statische deel van de uitleg (meestal fysieke kenmerken van het onderwerp), de tweede lijn het meer dynamische deel (vooral de bediening van een apparaat of het verloop van gebeurtenissen). Deze schema's vertonen veel overeenkomst met de '*declarative*' en '*procedural strategies*' van Paris en McKeown (1987) en de strategieën voor '*descriptive*' en '*sequence framing*' van Witte en Chery (1986).

Figuur 1 geeft het PISA-resultaat bij toepassing op de Sinterklaas-uitleg uit Tabel 1. Om de presentatie niet onnodig te compliceren, beperken we ons hier tot de hiërarchische structuur (voor volledig uitgewerkte voorbeelden, zie Van der Pool, 1995; Sanders & Van Wijk, 1996a). De structuur wordt van links naar rechts besproken, in dezelfde richting dus als waarin de tekst gelezen wordt. Het onderwerp wordt geïntroduceerd met een titel (s0) en een soort definitie (s1). In de volgende segmenten worden details besproken van een secundair personage, de Zwarte

Pieten (s2 to s9). Deze 'afdwaling' eindigt met een overstap naar de *Handelingslijn*: vanaf s10 richt de tekst zich op de viering van het Sinterklaasfeest. Eerst een beschrijving van wat er voorafgaat aan vijf december (s10 t/m s16), dan van de activiteiten op de dag zelf (s20 t/m s33). Deze uitleg wordt onderbroken door een korte terugkeer naar de *Kenmerklijn* waarin een nieuw aspect van de Sint wordt beschreven, zijn boek waarin alles staat over ons gedrag (s17 t/m s19). Met segment 33 komt de uitleg aan zijn eind (Sint is weg, de kinderen zijn blij: "het feest is voortbij"). Desondanks besluit de schrijver nog wat details aan de *Kenmerklijn* toe te voegen (s34 t/m s35a; let op de markering *Oh ja*). s36 is een meta-opmerking over de volledigheid van de tekst, wederom gevolgd door wat aanvullende opmerkingen over Sinterklaas (s37) en Zwarte Piet (s38).

Kortom, de globale organisatie van de tekst heeft te lijden van een zijspoor (s2 t/m s9), een onderbreking (s17 t/m s19) en wat losse eindstukjes (s34 t/m s38). Deze bevindingen laten de praktische waarde van een PISA-structuur zien als criterium voor tekstevaluatie en als richtsnoer voor tekstrevisie (Sanders & Van Wijk, 1991). De structuur geeft bijvoorbeeld aan welke segmenten 'er wat bij hangen' en op welke positie ze beter op hun plaats zouden zijn. Maar een PISA-structuur kan ook voor een meer theoretisch doel gebruikt worden: het identificeren van de productiestrategieën van een schrijver. Deze cognitieve pretentie gaat ervanuit dat in een PISA-structuur het 'schrijfplan' kan worden teruggevonden. De tekst in Tabel 1 geeft empirische ondersteuning voor deze aanname. Daarvoor kijken we naar tekstkenmerken die niet zoals de PISA-criteria syntactisch of lexicaal van aard zijn, namelijk de vier interpunctie-conventies van de schrijver: de punt en de '@' (d.w.z. begin op een nieuwe regel en zet de eerste letter in kapitaal) als markeringen in de globale structuur, de komma en de ronde haken als markeringen in de lokale structuur.

Figuur 1. PISA-structuur, interpunctie en referentiële continuïteit van de voorbeeldtekst.



Validering PISA-structuur

Figuur 1 presenteert onder het diagram twee typen aanvullende informatie. De eerste twee regels bevatten de codes voor interpunctie en pagina-opmaak: de komma, de punt, de ronde haakjes () en

het met een hoofdletter beginnen op een nieuwe regel (aangegeven met '@'). Geen van deze kenmerken speelt een rol in PISA. Daarom kunnen zij onafhankelijke evidentie bieden voor de plausibiliteit van de toegekende structuur. De volgende twee regels specificeren de referentiële continuïteit met betrekking tot tekstschema's, referenten en tijd. Een overgang van *Kenmerk-* naar *Handelingslijn* en vv. is aangegeven met een dubbele lijn (||), de wisseling van protagonist met een enkele lijn (|) en de naam van de betrokken partij (*Sint*, *Piet*, *kind*, *'zij'*) en nominale bepalingen van tijd met een enkele lijn (|) en het eerste woord van de frase (*als...*, *Op...*). Tijdens de analyse houdt PISA het verloop van deze typen continuïteit bij.

In de globale groepering van segmenten stemmen PISA en de schrijver sterk overeen. Van de 37 T-uniteitsgrenzen zijn er 16 met een @ gemarkeerd (43 procent). Referentiële overgangen in tekstschema, referent of tijd (aangegeven met || en |) scoren beduidend hoger: twaalf van de veertien werden gemarkeerd met een @ (86 procent; $\chi^2(1)=5.78$, $p<.02$). Het is zeer onwaarschijnlijk dat de schrijver door toeval haar @'s zo vaak op een referentiële overgang plaatste. Gegeven 37 T-uniteitsgrenzen, 14 referentiële overgangen en 16 @'s zou bij aselechte toewijzing gemiddeld genomen 6.05 keer een @ op een referentiële overgang terechtkomen. De kans dat dit door toeval 12 of meer keer gebeurt, is gelijk aan .0000645 (hypergeometrische kansverdeling voor steekproefrekking zonder teruglegging uit een eindige populatie). De schrijver startte op een nieuwe regel (= @) om een referentiële overgang te markeren, en geeft daarmee onafhankelijke steun aan de globale PISA-structuur.

Ten aanzien van de lokale onderschikking van segmenten stemmen PISA en de schrijver eveneens sterk overeen. Van 46 segmentgrenzen zijn er zestien gemarkeerd met @. Er resteren dus 30 kandidaten voor een andersoortige markering zoals komma, ronde haken of een combinatie van deze twee. Zo'n markering kwam voor op 12 van de 30 grenzen (40 procent). Segmenten die door PISA zijn geïdentificeerd als de definitie van een begrip dat in het direct voorafgaande segment werd geïntroduceerd, scoorden 11 uit 13 (85 procent; $\chi^2(1)=5.57$, $p<.02$). Een aselekt gebruik van beide markeringen zou leiden tot de markering van 5.20 definities. De kans dat dit door toeval elf keer of meer gebeurt, is één op het miljoen. De schrijver gebruikte de komma en ronde haakjes als markeringen van een terzijde waarin de betekenis van een bepaald begrip werd uitgelegd, en geeft daarmee steun aan een belangrijk aspect van de lokale PISA-structuur.

3. Stap 2: van structuur naar proces

De PISA-structuur geeft informatie over sterke en zwakke aspecten van de organisatie van informatie in een tekst. Maar ook is eruit af te leiden welk plan ten grondslag lag aan het schrijven van de tekst. De reconstructie van dit plan gaat uit van de twee eisen die volgens Bereiter en Scardamalia (1984), naast het bij je onderwerp blijven, gelden voor iedere serieuze poging een tekst te maken: de *lokale coherentie-eis* (segmenten moeten op een of andere wijze aansluiten op hun voorgangers) en de *niet te mager-eis* ('void filling'; een tekst moet aan bepaalde impliciete normen voldoen met betrekking tot bijvoorbeeld lengte en inhoud). Het plan wordt daarom in twee delen opgesplitst: een hoofdplan om topische en lokale coherentie te realiseren en een noodplan om de tekst niet te ondermaats te laten zijn.

De plannen die volgens onze analyse ten grondslag liggen aan de Sinterklaas-uitleg, staan in (1) en (2). Ieder plan bestaat uit een *intentie-strategie-procedures* combinatie (Van der Pool & Van Wijk, 1995). De intentie specificereert het probleem dat de schrijver moet oplossen; de strategie geeft in algemene termen de oplossing weer; de procedures omschrijven in detail de handelingen om de oplossing te realiseren. De precieze invulling van een plan moet voor iedere te analyseren tekst opnieuw worden bepaald (voor de reconstructie van een *Telefoon-uitleg*, zie Sanders & Van Wijk, 1996b). En zelfs als plannen gelijk lijken te zijn, kunnen schrijvers sterk van elkaar verschillen in hun feitelijk gebruik tijdens het schrijven; sommige schrijvers gebruiken bijvoorbeeld een noodplan veel terughoudender dan anderen.

- (1) **Hoofdplan**
 Intentie: Ik wil een uitleg geven van onderwerp X
 Strategie: Maak gebruik van relevante tekstschema's
 Procedures: *P1a* Beschrijf de bediening van X (=vul *Handelingslijn* in)
 P1b Geef bij termen die de lezer waarschijnlijk niet kent, een definitie of voorbeeld
- (2) **Noodplan**
 Intentie: Ik wil inhoud genereren over onderwerp X
 Strategie: Maak gebruik van trefwoorden in de tekst
 Procedures: *P2a* Griep terug op het onderwerp (=topic)
 P2b Griep terug op een lokaal begrip

Twee kenmerken van (1) en (2) verdienen vermelding. Ten eerste wordt de *Kenmerklijn* niet genoemd ofschoon deze wel deel uitmaakt van de PISA-structuur (zie Figuur 1). In de voorbeeldtekst was de *Kenmerklijn* niet een gepland onderdeel, maar een toevallig gevolg van de toepassing van eerst procedure *P1b* (de definitie in s1) en dan van herhaaldelijk *P2a* (de aangehechte laboraties in s17, s34 en s37). Ten tweede kunnen de procedures in (1) en (2) niet als recht-toe recht-aan instructies worden gezien. Bij toepassing activeert ieder van hen een subplan waarin zijzelf als intentie fungeren. Bijvoorbeeld, als *P1b* wordt uitgevoerd, wordt deze geïnterpreteerd als de 'sub'-intentie 'Ik wil een *term definiëren of illustreren*' die wordt gerealiseerd met de strategie 'Geef *aanvullende informatie*' en een van de procedures uit de reeks maak een analogie, een sematische decompositie, een stipulatie of een lijst voorbeelden. Om de discussie niet onnodig te compliceren hebben we in (1) en (2) alleen de toplaag van de plannen gegeven. Verder gespecificeerde voorbeelden geven wij in TRACE-modellen (een acroniem voor Tekst-gebaseerde Reconstructies van Activiteiten in de Conceptualisator). Zulke modellen zijn opgesteld voor persoonsbeschrijvingen (Van der Pool, 1995, deze bundel) en pro-contra argumentaties (Van Wijk, 1995).

Met behulp van de plannen (1) en (2) kan ieder segment van de PISA-structuur gezien worden als het resultaat van een van de vier procedures. Tabel 2 geeft de resultaten van deze reconstructie. Voor ieder segment is aangegeven welke procedure eraan ten grondslag ligt en het thema of de antecedent waarop gefocust wordt. De tabel geeft ook aan waar de schrijfter van type procedure

verandert, oftewel waar de belangrijkste veranderingen in de structuur plaats hebben. De coderingen in de tabel zijn gelijk aan de conventies in de onderste regels van Figuur 1.

- Iedere keer als (een reeks) *P1a*, *P2a* of *P2b* begint, wordt dit aangegeven met een dubbele onderstreping (*P1b* speelt geen rol: deze procedure heeft een lokale functie en onderbreekt niet de rode draad van de tekst);
- Binnen een reeks van *P1a*-aanroepen geeft een enkele onderstreping aan (i) een wisseling van protagonist (de nieuwe staat vet gedrukt) en (ii) een wisseling van tijd door een nominale omschrijving (zoals in s16a en s29a, segmentnummer staat vet gedrukt).

Tabel 2. *Reconstructie van de PISA-structuur in termen van de mentale plannen (voor iedere procedure is thema/antecedent gegeven).*

segment	procedure	segment	procedure	segment	procedure
0	--- sint	12	P1a piet	27	P1a piet
1	<u>P1b sint</u>	13	<u>P1b schoorsteen</u>	28	<u>29a- P1a 'ze'</u>
1a	<u>P1b boot</u>	14-15	<u>P1a kind</u>	29-30	<u>P1a 'ze'</u>
2-3	<u>P1b piet</u>	16a-16	<u>P1a kind</u>	31-32	<u>P1a sint</u>
4	<u>P1b piet</u>	16b	<u>P1b bed</u>	33-34	<u>P1b Spanje</u>
5	<u>P1b roe</u>	17	<u>P2a sint</u>	35-35a 36	<u>P1a kind</u>
6	<u>P1b piet</u>	18-19	<u>P1b boek</u>	37	<u>P1b blij</u>
7a-8	<u>P1b zak</u>	20	<u>P1a sint</u>	38	<u>P2a sint</u>
7b	<u>P1b kind</u>	21	<u>P1b 5 december</u>		<u>P1b paard</u>
7c	<u>P1b stout</u>	22a-23	<u>P1b-jarig zijn</u>		---
9	<u>P1b cadeautje</u>	24	<u>P1a piet</u>		<u>P2a sint</u>
10	<u>P1a piet</u>	25	<u>P1a piet</u>		<u>P2b piet</u>
11	<u>P1b huis</u>	26	<u>P1b deur</u>		---

Validering schrijfflannen

In totaal bevat de tekst tien zelfcorrecties en negentien spel- en taalfouten (waarvan er zes spontaan zijn gecorrigeerd). De verdeling ervan is verre van willekeurig. Als we de tekst in tweeën delen (tussen segment 19 en 20), blijkt dat de eerste helft 0.08 fouten per segment bevat, en de tweede helft 0.77 fouten ($\chi^2(1)=12.230, p<.001$). De eerste helft bevat 0.04 zelf-correcties per segment, de tweede helft 0.41 ($\chi^2(1)=5.86, p<.02$). Met frequentie neemt ook diversiteit toe. In de eerste helft treffen we alleen een spelfout (SP) en een vorm-correctie (***) aan; in de tweede helft ook taalfouten (SY...) en inhoud-revisies (**), [***]. Een oppervlakkige interpretatie van deze bevindingen zou zijn "hoe langer je bezig bent, des te groter de kans op ontsporingen". Maar waarom zou dat zo zijn? Komt het door vermoeidheid, verminderde motivatie, aandachtsvlies? Mogelijk, maar de invloed van deze factoren blijkt ook niet los gezien te kunnen worden van de schrijfflannen (1) en (2). In Tabel 2 is te zien hoe ieder segment het resultaat was van een procedure. Tabel 3 geeft voor iedere procedure aan hoe hun segmenten zich gedroegen in termen van fouten, vorm-correcties en inhoud-revisies (de correcties bevatten geen taalfouten omdat deze nooit gecorrigeerd zijn).

Tabel 3. *Fouten en zelfcorrecties in relatie tot procedure en positie in tekst.*

volgnummer segment	procedure	fouten per segment	percentage herstelde spelfouten	inhoud revisies per segment	aantal segmenten
	<i>P1b</i>	.11	100	.00	27
	<i>P2a</i>	.67	100	.00	3
	<i>P1a</i>	.73	33	.27	15
	<i>P2b</i>	3.00	50	.00	1
1-9	<i>P1b</i>	.00	-	.00	13
10-37	<i>P1b</i>	.21	100	.00	14
10-16	<i>P1a</i>	.33	100	.00	6
20-33	<i>P1a</i>	1.00	50	.44	9
			25		

Wanneer de procedures groepsgewijs worden bekeken, vallen drie verschillen op (Tabel 3, bovenste regels). (i) Het aantal fouten per segment varieerde van .11 voor *P1b* tot 3.00 voor *P2b* met voor *P2a* en *P1a* tussenvallende waarden ($\chi^2(3)=18.84, p<.001$). (ii) Het percentage gecorrigeerde spelfouten vertoonde een trend in tegengestelde richting: het hoogst bij *P1b* en *P2a*, het laagst bij *P1a* en *P2b*. (iii) Inhoud-revisies traden alleen bij *P1a* op.

Bij aparte beschouwing blijken de procedures zich verschillend te gedragen met betrekking tot positie in de tekst (Tabel 3, onderste regels). Gezien het kleine aantal waarnemingen beperken we ons tot de procedures van het hoofdplan. Positie in de tekst heeft weinig invloed op *P1b*. In vergelijking met de eerste aanroepen (s1-9) vertonen de latere gevallen (s10-37) meer fouten, maar die worden, wat spelling betreft, voor 100 procent spontaan gecorrigeerd. De procedure *P1a* lijkt wel beïnvloed te worden door positie. Latere aanroepen (s20-33 versus s10-16) laten niet alleen een forse toename van fouten zien, maar ook een afname van 'monitoring efficiency': er worden minder spelfouten gecorrigeerd.

De procedures verschillen in de frequentie van fouten, zelf-correcties en inhoud-revisies, en in gevoeligheid voor positie in de tekst. Deze bevindingen suggereren dat haperingen in tekstproductie en 'monitoring' niet alleen bepaald worden door motivatie of vermoeidheid, maar ook door factoren die intrinsiek behoren tot de procedures. In de volgende paragraaf stellen wij een cognitief model voor ter verklaring van deze factoren.

4. Enige theoretische overdenkingen

De uitvoering van een schrijfprocedure vergt een bepaalde mentale last. Deze last kan twee gedaanten aannemen: als capaciteitseisen aan het werkgeheugen of als inspanning om informatie op te halen uit het semantisch/episodisch geheugen. De eerste factor, geheugenbeperkingen, wordt van belang als "the writer aspires to something beyond minimally adequate text" (Bereiter &

Scardamalia, 1984, p.424), als de schrijver besluit om inhoud grondiger te bewerken en retorische overwegingen toe te passen (voor een gedetailleerde uitwerking van de rol van geheugenbeperkingen bij het genereren van een pro-contra argumentatie, zie Van Wijk, 1995). De tweede factor, activatie van het geheugen, is van belang wanneer inhoudsvinding de centrale taak is en de tekststructuur of coherentie als een bij-product tot stand komt eerder dan als een bewust geplande constructie, precies de werkwijze dus die aan de voorbeeldtekst ten grondslag lag.

Om te bepalen hoe gemakkelijk de procedures in (1) en (2) kunnen worden uitgevoerd, stellen wij voor dat het ophalen van informatie afhangt van drie kenmerken van het antecedent of thema: afstand, aansluiting en bepaaldheid. Deze drie zijn geordend: afstand gaat gepaard met de minste last, bepaaldheid met de meeste. Dit komt door verschillen in de aard van het uit te voeren mentale proces: bij afstand is dat directe activatie, bij aansluiting komt daar filteren bij en bij bepaaldheid is een geheel nieuw proces nodig, namelijk gericht zoeken (*focussing*, zie ook Schilperoord, 1996).

Afstand: hoe verder weg een antecedent, des te meer inspanning. Dit maakt *P1b* de gemakkelijkste van de vier: er moet immers informatie gevonden worden voor een concept uit het laatst geproduceerde segment en dus met de hoogste graad van activatie (in ieder geval meer dan het nogal diffuse topic aangekaart door *P2a*).

Aansluiting: hoe meer je moet letten op wat al geschreven is, des te meer inspanning. Dit maakt *P2a* de op een na gemakkelijkste; je hoeft alleen maar te voorkomen hetzelfde twee keer te noemen terwijl *P1a* ook verlangt dat de informatie in een bepaald temporeel patroon van handelingen past.

Bepaaldheid: hoe meer onzekerheid over de keuze van het thema waarop gereageerd moet worden, des te meer inspanning. Dit zet *P1a* op de derde plaats waarna *P2b* overblijft als de moeilijkst uit te voeren procedure.

Samenvattend, de ordening van de procedures in termen van hun mentale belasting is: $P1b < P2a < P1a < P2b$, waarbij '<' staat voor 'is gemakkelijker dan'.

De resultaten in tabel 3 sluiten aan bij deze geheugen-activatie verklaring. De voorgestelde ordening van procedures komt terug in de data: als werklust toeneemt, neemt foutieve productie ook toe en gaat 'monitoring' activiteit naar beneden. De factor *Aansluiting*, dat wil zeggen de eis om ook te letten op wat al geschreven is, is het meest van toepassing op *P1a*. Alleen deze procedure scoort op inhoud-revisies. De factor *Afstand*, dat wil zeggen de mate van activatie op een bepaald moment, is het meest in het voordeel van *P1b*. Deze procedure scoort het laagst op fouten en de efficiëntie ervan staat grotendeels los van positie in de tekst.

5. Discussie

Bij de besproken casus springen twee resultaten in het oog. Ten eerste, het zou niet eerlijk zijn tegenover de schrijfter om haar interpunctie-conventies gebrekkig of zelfs fout te noemen. Zij blijkt gewoon een andere gebruiksregel te hanteren, namelijk een die gebaseerd is op tekststructurele overwegingen en niet op zinsbouw. Ten tweede, haar taal- en spelfouten moeten

als procesgegevens behandeld worden, niet als blijken van een gebrekkige regelkennis. Hiermee wordt tevens verklaard waarom zij haar spelfouten zo inconsequent maakte (bijvoorbeeld *zetten/zeten* in s14 en s27; *heeft/heef* in s37 en s38). Belangrijker dan deze ene casus is natuurlijk de vraag wat tekstanalyse als onderzoeksparadigma waard is. We gaan hier dieper op in eerst voor de betrouwbaarheid, dan voor de validiteit.

Objectiviteit van de tekstanalyse is een belangrijk motief geweest om PISA te ontwikkelen. Nu deze doelstelling gerealiseerd is (Van der Pool, 1995; Sanders & Van Wijk, 1996a) is onze grootste zorg de generaliseerbaarheid van de bevindingen.¹ Voor de hier besproken tekst hebben we een relatie vastgesteld tussen tekststructuur, linguïstische markeerders, interpunctie, taal- en spelfouten en zelf-correcties. We hebben ook de mentale plannen kunnen reconstrueren. Maar hoe ideosyncratisch zijn deze bevindingen? Gedragen andere schrijvers zich op dezelfde wijze of is dan een veel grotere diversiteit te vinden? Slechts een weg biedt antwoord op deze vragen: de analyse van grote tekstbestanden, bij voorkeur van schrijvers van verschillend vaardigheidsniveau die over verschillende onderwerpen hebben geschreven in reactie op verschillende instructies. Voor iedere tekst in zo'n bestand moet dan een verslag gemaakt worden zoals in dit paper. Door een vergelijking van PISA-structuren en de daaruit afgeleide schrijfplannen kunnen we dan conclusies trekken over de onderliggende mentale processen en de ontwikkeling daarvan (voor een eerste voorbeeld van zo'n grootschalig project, zie Van der Pool, 1995).

De vraag naar de validiteit vereist een geheel andere aanpak. Voor de hier besproken teksten we de validiteit nagaan aan de hand van kenmerken die allemaal aan de tekst zelf ontleend waren. Maar naarmate een schrijver vaardiger wordt, past deze interpunctietekens steeds meer geconventionaliseerd toe en verdwijnen de meeste fouten. Dan moeten we wel naar samenhang gaan zoeken met kenmerken buiten de tekst. Een veelbelovende weg naar tekst-externe validering is het onderzoek van temporele patronen in het schrijfproces (Sanders et al., 1996). Pauze-gedrag blijkt nauw gerelateerd te zijn aan de structurele eigenschappen van het te produceren segment (Schilperoord, 1996; Schilperoord & Sanders, 1997). Een combinatie van tekstanalyse en tijdregistratie volstaat echter niet voor validiteit. De vraag naar volledigheid moet namelijk ook gesteld worden. Wanneer schrijvers vaardiger worden, reflecteert het schrijfproces zich niet alleen in wat wel in de tekst staat, maar ook in wat er juist niet in staat. Hoe heeft een schrijver de intentie uitgewerkt alvorens met schrijven te beginnen? Welke ideeën zijn opgekomen, maar uiteindelijk niet in de tekst terecht gekomen? Om toegang te krijgen tot deze processen zijn methodes nodig zoals hardop denken, retrospectieve verslagen en aantekeningen maken (Van den Bergh & Rijlaarsdam, 1996).

Samenvattend, schrijfprocesonderzoek moet een combinatie zijn van tekstanalyse, pauzemetingen en verslagtechnieken. Deze combinatie van 'on-line' en 'off-line' technieken zal het meest vruchtbaar blijken te zijn. We hopen met ons werk ertoe bij te dragen dat tekstanalyse een volledig geïntegreerd lid wordt van de basismethoden in het schrijfprocesonderzoek.

Generaliseerbaarheid wordt onder validiteit geplaatst wanneer in de vergelijking één van de gegevensbestanden als norm wordt beschouwd ('cross validation') en onder betrouwbaarheid wanneer het gaat om de herhaalbaarheid van observaties ('replication'). Wij hanteren het begrip in de tweede betekenis; het gaat ons om het exploreren van mogelijkheden en (nog) niet om het toetsen van eerdere bevindingen. Een strikte scheiding tussen validiteit en betrouwbaarheid is overigens moeilijk te maken (Cronbach, Rajaratnam & Gleser, 1963).

Literatuur

- Bergh, H. van den & Rijlaarsdam, G., (1996), The dynamics of composing: Modeling writing process data. In: M. Levy & S. Ransdell (Eds.), *The Science of Writing* (207-232). Mahwah NJ: Erlbaum.
- Bereiter, C. & Scardamalia, M., (1983), Levels of inquiry in writing research. In: P. Mosenenthal, L. Tamor & S. Walmsley (Eds.), *Research on Writing: Principles and Methods* (3-25). New York: Longman.
- Bereiter, C. & Scardamalia, M., (1984), Information-processing demand of text composition. In: H. Mandl, N. Stein & T. Trabasso (Eds.), *Learning and Comprehension of Text* (407-428). Hillsdale NJ: Erlbaum.
- Cooper, C. & Matsuhashi, A., (1983), A theory of the writing process. In: Martlew, M., (Ed.), *The Psychology of Written Language* (3-39). Chichester: Wiley.
- Cronbach, L., Rajaratnam, N. & Gleser, G., (1963), Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16, 137-163.
- Paris, C. & McKeown, K., (1987), Discourse strategies for describing complex physical objects. In: G. Kempen (Ed.), *Natural Language Generation* (97-115). Dordrecht: Nijhoff.
- Pool, E. van der (1995), *Writing as a conceptual process: A text-analytical study of developmental aspects*. Proefschrift KU Brabant.
- Pool, E. van der & Wijk, C. van, (1995), Proces en strategie in een psycholinguïstisch model van schrijven en lezen. *Tijdschrift voor Onderwijsresearch*, 20, 200-214.
- Sanders, T. & Wijk, C. van, (1991), Tekststructuur en tekstvaliditeit: van analyse naar evaluatie en revisie. In: W. Vuijk & M. Bax (red.), *Thema's in de Taalbeheersing* (140-150). Dordrecht: ICG.
- Sanders, T. & Wijk, C. van, (1994), Procedures voor Incrementele Structuur-Analyse (PISA): Uitleg en toepassing op verklarende teksten. In A. Maes, P. van Hauwermeiren & L. van Waes (red.), *Perspectieven in Taalbeheersingsonderzoek* (191-203). Dordrecht: ICG.
- Sanders, T. & Wijk, C. van, (1996a), PISA - A procedure for analyzing the structure of explanatory texts. *Text*, 16, 91-132.
- Sanders, T. & Wijk, C. van, (1996b), Text analysis as a research tool. In: M. Levy & S. Ransdell (Eds.), *The Science of Writing* (251-269). Mahwah NJ: Erlbaum.
- Sanders, T., Janssen, D., Pool, E. van der, Schilperoord, J., & Wijk, C. van, (1996), Hierarchical text structure in written products and writing processes. In: G. Rijlaarsdam, H. van den Bergh & M. Couzijn (Eds.), *Theories, Models and Methodology in Writing Research* (473-492). Amsterdam: Amsterdam University Press.
- Schilperoord, J., (1996), *It's about time: Temporal aspects of cognitive processes in text production*. Proefschrift, Universiteit Utrecht.
- Schilperoord, J. & Sanders, T., (1997), Pauses, cognitive rhythms and discourse structure: An empirical study of discourse production. In: A. Liebert & G. Redeker (Eds.), *Cognitive Linguistics at the Discourse Level*. Amsterdam: Benjamins.
- Wijk, C. van & Sanders, T., (1991), De rol van een standaard-tekstmodel bij de analyse van spontaan geschreven teksten. In: W. Vuijk & M. Bax (red.), *Thema's in de Taalbeheersing* (131-139). Dordrecht: ICG.
- Wijk, C. van, (1992), Information analysis in written discourse. In: J. de Jong & L. Verhoeven (Eds.), *The Construct of Language Proficiency* (85-99). Amsterdam: Benjamins.
- Wijk, C. van, (1995, May), *Levels of competence in writing*. Paper presented at the 5th European Workshop on Natural Language Generation, Leiden, The Netherlands.

Els van der Pool

1. De schrijver aan het werk

Hoe zal ik mijn tekst beginnen? Welke informatie neem ik wel op, welke toch maar niet? In welke volgorde kan ik het beste mijn mededelingen doen? Allemaal vragen die een schrijver bezig houden bij het maken van een tekst. Naast het kiezen van de juiste woorden en het formuleren van zinnen houdt een schrijver zich bezig met het bedenken van en denken over de informatie die hij (al dan niet) in zijn tekst opneemt (Bereiter & Scardamalia, 1987; Flower & Hayes, 1981; Levelt, 1989¹). Dit conceptuele deel van het schrijfproces bestaat grofweg uit *Reflecteren* op en *Selecteren* van informatie.

Bij het *Reflecteren* overweegt de schrijver zijn doel dat hij wil bereiken met zijn tekst (wil hij bijvoorbeeld overtuigen of amuseren) en bepaalt een strategie om dat doel te realiseren, waarbij ook het beeld dat de schrijver heeft van zijn lezer een rol speelt (bijvoorbeeld welke voorkennis heeft de lezer). Bij het *Selecteren* bedenkt de schrijver welke informatie al dan niet geschikt is om op te nemen, en bepaalt een volgorde voor de informatie die hij wil opnemen. Afhankelijk van zijn intentie, bijvoorbeeld amuseren, zal de schrijver bepaalde informatie - grappige- wel en andere -serieuze- niet opnemen. Ook de plaats van de informatie in de tekst bepaalt de schrijver aan de hand van zijn doel; wil hij een lezer bijvoorbeeld overtuigen, dan zal hij beginnen met het sterkste argument.

De precieze beslissingen die een schrijver neemt bij het Reflecteren en Selecteren zijn minder bekend. Waarin verschillen schrijvers van elkaar? Wat doet een ervaren schrijver anders dan een 10-jarige beginner? Waaruit blijkt dat in de tekst? Relevante vragen die een antwoord behoeven, niet alleen om meer zicht te krijgen op menselijk taalgedrag, maar ook om te weten wat die beginner moet worden geleerd om hem vorderingen te laten maken. In deze bijdrage staan twee vragen centraal:

- (a) hoe kunnen we conceptuele processen bij het schrijven expliciteren, in regels beschrijven?
- (b) hoe ontwikkelen deze processen zich met leeftijd, van 10-jarige beginnende schrijvers tot volwassen schrijvers?

Bij het beantwoorden van deze vragen gaan we ervan uit dat teksten sporen bevatten van conceptuele delen van schrijven. Welk spoor we precies volgen en hoe, is beschreven in paragraaf 2. In paragraaf 3 staat beschreven, in antwoord op de eerste vraag, hoe we gevonden sporen gebruiken om stappen die een schrijver heeft gezet te reconstrueren en te expliciteren. In paragraaf 4 staan, in antwoord op de tweede vraag, de resultaten die laten zien hoe die conceptuele