

Joost Schilperoord

Inleiding

Zoals ieder uit eigen ondervinding weet, gaat het produceren van taal in horten en stoten. Dit verschijnsel wordt aangeduid met de term *incrementaliteit* (Levelt 1989), het idee dat taaluitingen niet zozeer als geheel gepland en vervolgens uitgesproken worden, maar dat inhoud en vorm als het ware aangroeitend, dus per increment, tot stand komen. Een vraag die de psycholinguïstiek reeds lange tijd bezighoudt is of er systematiek aan die incrementen ten grondslag ligt, en zo ja welke. Het idee is, algemeen gesteld, dat er een bepaalde *eenheid* van taalproductie is. Die verwerkingseenheid zou dan moeten corresponderen met een bepaalde lexicale of grammaticale structuur in de spraakuiting. De analyse van pauze patronen in spraak kan voor die vraag relevante informatie opleveren, zoals bondig verwoord is door Boomer (1965):

'(...) *hesitations in spontaneous speech occur at points where decisions and choices are being made. On this basis, the patterning of hesitations should provide clues as to the size and nature of the encoding units which are operative.*' (Boomer 1965, 148)

Het behoeft geen betoog dat inzicht in de eenheden die zijn betrokken in de productie van taal van groot belang is voor de ontwikkeling van een psycholinguïstische theorie van taalproductie. Met betrekking tot de interface tussen het conceptuele niveau van de 'boodschap' en het talige niveau van de 'taaluiting', is, met name op basis van observaties aan pauze patronen in spraak, de (grammaticale) *clause* voorgesteld als codeereenheid (o.a. in Boomer 1965, Ford & Holmes 1978, Ford 1982, Holmes 1988, Levelt 1989). Hoewel de gehanteerde clause-definities in deze studies nogal verschillen, houdt dit voorstel dus in zijn algemeenheid in dat de clause als linguïstische structuur correspondeert met de conceptuele structuur die het conceptuele systeem als 'eenheid' prepareert en aanlevert.

Psycholinguïsten onderzoeken echter doorgaans (spontane) mondelinge taalproductie. De vraag die in dit artikel centraal staat is of de clause eveneens de eenheid is van *schriftelijke tekstproductie*. Om die vraag te onderzoeken is een analyse uitgevoerd van pauze patronen die zich voordoen bij de productie van korte, routineuze zakenbrieven. Daartoe werd een corpus aangelegd van 120 dicteerprocessen van zes professionele advocaten. Hoewel dicteren niet een voor de hand liggende modus van taalproductie is, kan het toch opgevat worden als een vorm van schriftelijke tekstproductie. Dictafonie wordt gebruikt om brieven, memo's, verzoeken e.d. op te stellen, wat allemaal typische schriftelijke teksten zijn. Bovendien zijn dicteerders zich goed bewust van het feit dat ze schriftelijke teksten opstellen, hetgeen blijkt uit zaken als het interpungeren van teksten, het inspreken van lay-out instructies, en uit het feit dat bijvoorbeeld reparaties in teksten altijd *over de oorspronkelijke delen* heen worden ingesproken. Zulke

- Sanders, T. & Wijk, C. van, (1996a), Text analysis as a research tool: how hierarchical text structure contributes to the understanding of conceptual processes in writing. In M. Levy & S. Ransdell (Red.), *The science of writing* (251-269). Mahwah NJ: Erlbaum.
- Sanders, T. & Wijk, C. van, (1996b), PISA - A procedure for analyzing the structure of explanatory texts. *Text* 16, 91-132.
- Steehouder, M., Jansen, C., Maat, K., Staak, J. van der & Woudstra, E., (1992), *Leren Communiceren. Handboek voor mondelinge en schriftelijke communicatie*. Groningen: Wolters-Noordhoff.
- Witte, S. & Cherry, R., (1986), Writing processes and written products in composition research. In C. Cooper & S. Greenbaum (Red.), *Studying writing: linguistic approaches* (112-153). Beverly Hills: Sage.
- Wijk, van C., (1992), Information analysis in written discourse. In J. de Jong & L. Verhoeven (Red.), *The construct of language proficiency* (85-99). Amsterdam: Benjamins.
- Wijk, van C., (1995), *Levels of competence in writing*. Paper gepresenteerd op de 5th European Workshop on Natural Language Generation in Leiden.

afwijkingen van het strikt lineair-temporeel bepaalde proces van *mondeling* taalgebruik zijn juist typerend voor schriftelijke taalproductie (zie ook: Schilperoord 1996, 19-23).

Drie typen analyses van pauzeerpatronen zijn uitgevoerd. Allereerst is gekeken naar de pauzeerparameters (geproportioneerde) *frequentie* en *lengte* van pauzes op verschillende typen grammaticale en tekstuele locaties. Daarnaast is een analyse uitgevoerd van de ontwikkeling van pauze-lengtes *binnen* bepaalde structuren. Hierbij gaat het om de vraag hoe pauzeerlengte zich ontwikkelt als functie van de plaats van pauzes binnen, bijvoorbeeld, een clause of een zin (na het eerste woord, na het tweede etc.). Een nadere motivering voor dit type analyse volgt in paragraaf 3.

Dit artikel is als volgt opgebouwd. Paragraaf 1 biedt een nadere rechtvaardiging voor een aparte studie naar eenheden in *schriftelijke* taalproductie. Ik zal, op basis van een bespreking van twee transcripten van mondelinge en schriftelijke taalproductie, laten zien dat er nogal wat verschillen bestaan in clause-structuren in beide modaliteiten. Paragraaf 2 rapporteert over de uitkomsten van een analyse waarin de samenhang centraal staat tussen pauze-locaties enerzijds, en de geobserveerde frequenties en lengte van pauzes op die locaties anderzijds. De uitkomsten worden in perspectief geplaatst door ze te vergelijken met de data van Ford (1982) met betrekking tot mondelinge spraak. De resultaten bieden steun aan de gedachte dat clauses ook in schriftelijke productie de primaire eenheid zijn. Additionele steun daarvoor biedt paragraaf 3, waarin de temporele modificatie-data worden gerapporteerd. Paragraaf 4 tenslotte bevat enkele conclusies.

1. Clausestructuren en relaties in gesproken en geschreven taal

Onderstaand transcript (1) van gesproken taalproductie is afkomstig uit een televisie-interview met Marco van Basten. Van Basten gaat in op de vraag van de interviewer hoe het was om Johan Cruyff als trainer te hebben. Het antwoord bestond uit de volgende monoloog. De transcriptregels zijn onderscheiden op basis van pauzes, aangeduid met een /-teken. Dubbele punten duiden klinker-verlengingen aan; (.) een korte, maar nog duidelijk waarneembare pauze, en (..) en (...): duiden langere pauzes aan.

- (1) 1. het was heel leerzaam ja / (...)
2. wij hebben toen (eh) / (...)
3. een aantal trainers gehad met Lindner en de Mos / (...)
4. enne: toen kwam Cruyff / (...)
5. en Spitz Kohn enzo / (...)
6. en toen kwam Cruyff / (...)
7. en dat was voor / (...)
8. een jonge groep die wij hadden was dat heel interessant en heel erg goed / (...)
9. en wij hebben heel veel geleerd van 'm / (...)
10. heel veel op tactisch gebied / (...)
11. enne: ja dat is op een gegeven moment toch / (...)
12. heel belangrijk / (...)
13. en je ziet ook in eh: / (...)
14. die: / (...)

15. in die tijd speelden de meeste spelers ook goed op niveau / (...)
16. zoals Rijkard toch ook op jonge leeftijd en Winter en Bosman / (...)
17. die zijn later allemaal weggegaan / (...)
18. maar als ze aan die periode terugdenken is het een hele fijne periode geweest d'r was een goede sfeer / (...)
19. en je speelde goed / (...)
20. en je was gewoon in een proces je was gewoon / (...)
21. (eh ja:) aan het studeren is overdreven maar we waren aan het leren / (...)
22. en Johan was daar de perfecte man in / (...)

Uitzonderingen daargelaten produceert Van Basten zijn verhaal in de vorm van incrementen die vrijwel samenvallen met simpele monoclausale constructies (bv. regels 1, 4, 6, 9, 13, 15, 17, 18, 19, 20 en 22) of minimaal samengetrokken clause (5, 10, 16). Zelfs een oppervlakkige beschouwing van (1) maakt dus duidelijk dat clauses een voor de hand liggende eenheidskandidaat zijn. De clauses vertonen nauwelijks enige inbedding, maar worden met vooral de conjunctie *en* aan elkaar geketend. Bovendien is er zo goed als geen sprake van inhoudelijk integratie - de clauses lopen geheel in de pas met Chafe's *one-new-concept-at-a-time* constraint (Chafe 1987). Met andere woorden, de spreker voegt per increment één nieuw inhoudselement toe aan de monoloog, en de meest voor de hand liggende structuur om dat te doen is de clause.

Deze proces- en productkenmerken contrasteren krachtig met de gang van zaken bij *schriftelijke* productie, zoals gedemonstreerd in transcript (2). De /-tekens duiden pauzes aan, waarvan de lengte in tienden van seconden is weergegeven.

- (2) [...]
1. Tijdens de / 4
2. bespreking die / 7
3. ik op / 5
4. DATUM jongstleden / 13
5. tezamen met de heer X / 3
6. van cliënte met / 19
7. de heer Y had / 29
8. heeft / 3
9. de heer Y / 54
10. voorgesteld een / 72
11. bedrag te betalen door cliënte / 60
12. tegen / 5
13. finale kwijting wegens / 9
14. schadevergoeding / 51
15. terwyl / 50
16. een / 11
17. bedrag van BEDRAG / 6
18. dat erkend / 29
19. door / 10
20. de heer / 4
21. Y aan cliënte verschuldigd is / 42
22. wordt / 6
23. kwijgescholden.
- [...]

Het zal zelfs Nederlandse taalgebruikers moeite kosten om in één oogopslag te doorzien hoe de hier geproduceerde zin in elkaar steekt. Toch zijn dit type constructies in mijn corpus bepaald geen uitzondering. Ik som kort enkele in het oog springende proces- en productverschillen met (1)

op. Allereerst bestaan de incrementen voornamelijk uit woorden en constituenten, waarbij de pauzes vaak juist niet met constituentengrenzen samenvallen (zie ook: Schilperoord 1997, Schilperoord & Verhagen 1997a). Ten tweede: waar fragment (1) in elk geval 12 hoofdzinnen bevat, bevat (2) er slechts één, die samengesteld is uit 5 clauses. Verder is er in (2) in plaats van clause-ketening eerder sprake van diepgaande clause-*inbedding*. Dit verschil in de structuur is weergegeven in (3). Hierbij betekent bbc: beperkende relatieve clause; ubc: uitbreidende relatieve clause; ac: adverbiale clause; bc: beknoppte clause; cc: complement clause; en hz: hoofdzin.

- (3) a. (...)_{hz}& (...)_{hz}& (...)_{bb}& (...)_{bb}& (...)_{bc}& (...)_{bc}& (...)_{bc}& (...)_{bc}& (...)_{bc}& (...)_{bc}& (...)_{bc}& (...)_{bc}& (...)_{bc}& (...)_{bc}& (...)_{bc}
- b. (...)_{bc}& (...)_{bc}

Tenslotte blijkt er (2) geen sprake van 'one-new-concept-at-a-time': typerend voor de incrementen lijkt juist dat ze fragmenten van deelconcepten uitdrukken waarvan de afronding/completering over meerdere incrementen plaatsvindt. De clauses in (2) zijn inhoudelijk beschouwd vergaand geïntegreerd. Dit is te zien in (4), waarin de clause-structuur en incrementen samen zijn weergegeven.

- (4) 1 2 2 7 8 11 11 11 12 14
 ((Tijdens... bespreking (die...had)) heeft... bedrag (te...cliënt) tegen... schadevergoeding
- 15 17 18 21 22 23
 (terwijl...BEDRAG (dat...is) wordt kwijtgescholden))

De aangestipte proces- als productverschillen tussen gesproken en geschreven clause-complexen kunnen aanleiding zijn tot drie opponerende verwachtingen aangaande de productie-eenheid bij schriftelijke taalproductie. Denkbaar is allereerst dat de productie-eenheid bij schriftelijke productie *kleiner* is, en dat die correspondeert met talige eenheden als woorden en constituenten. Daarop zou de aard van de productie-incrementen kunnen wijzen. Denkbaar is echter ook dat de productie-eenheid zich juist manifesteert in eenheden groter dan de clause - in (vol-)zinnen dus. Daarop zou het hoge integratie- en inbeddingsgehalte van de clauses in (2) kunnen wijzen. En tenslotte zou men kunnen betogen dat de eenheid van productie invariant is ten opzichte van de productiemodus, met andere woorden dat de omvang van de incrementen een artefact is van een veel 'bedachtzamer' productiewijze, en de integratie-grad van de structuur als geheel van het genre (de formele zakenbrief), maar dat deze factoren niet wijzen op een wezenlijk veranderde procesgang.

Hoe het zij, de geobserveerde verschillen rechtvaardigen een apart onderzoek naar de vraag welke eenheden van productie operatief zijn in schriftelijke taalproductie. De volgende twee paragrafen rapporteren de uitkomsten van een grootschalige pauze-analyse waarmee getracht wordt op die vraag een antwoord te geven.

2. Proportie en lengte van pauzes in mondelinge en schriftelijke taalproductie

De analyse van de temporele parameters 'proportie pauzes' en 'lengte van pauzes' vormt een

belangrijke bron van evidentie voor de betrokken eenheden in taalproductie. De techniek is gebaseerd op de gedachte dat wanneer een bepaalde talige eenheid X correspondeert met de mentale plan- en codeereenheid, dit geïntensiveerde pauzeactiviteiten uit zal lokken op de grenzen tussen opeenvolgende instanties van X. Concreet zou zich dat manifesteren als relatief lange pauzes, en relatief hoge pauzeerproporties op die plaatsen. In deze paragraaf vermeld en bespreek ik de resultaten van een frequentie- en lengte-analyse van pauzes in schriftelijke tekstproductieprocessen. Om de gegevens in perspectief te kunnen plaatsen, vergelijk ik ze met die uit een studie van Ford (1982) waarin een vergelijkbare analyse wordt gerapporteerd naar pauzeactiviteiten in *mondellinge* taalproductie.

Van alle geobserveerde pauzes in het corpus werd de lengte gemeten. De locatie van een pauze werd bepaald op basis van vijf categorieën: tussen alinea's, tussen zinnen, tussen clauses, tussen constituenten en tussen woorden. De categorieën zijn inclusief in de zin dat een hogere categorie altijd alle lagere impliceert.

Omdat het in dit artikel primair om de productie-status van clauses gaat, is het nodig om het gehanteerde clause-begrip toe te lichten. In het algemeen geldt als een clause iedere structuur die op syntactische (oppervlakte-)kenmerken als zodanig aangeduid kan worden. Zinnen (5) bevat derhalve twee clauses.

- (5) a. Wij hebben ons daartegen willen verzekeren / door de mogelijkheid van een aparte toegangscontrole op onze plaatsen te bedingen.
- b. Op dit moment vervult hij deze functie / doch een apart besluit zou zijn taken vergemakkelijken.

De voornaamste uitzonderingen op deze regel betreffen complement-clauses, beperkende relatieve clauses en graad-aanduidende clauses. Voorbeelden zijn gegeven in (6).

- (6) a. Daarbij komt dat zijn vrouw ernstig gehandicapt is (...)
- b. Hierbij zend ik een kopie van de telefax die ik ontving van (...)
- c. De situatie is echter in zoverre verschillend dat in PLAATSAAM reeds voor DATUM sprake was van *accu-opslag* (...)

De voornaamste reden voor deze uitzonderingen was dat gebeurtenis of toestand die in een (op syntactische gronden onderscheiden) clause wordt uitgedrukt op zichzelf conceptualiseerbaar diende te zijn. Daarvan is sprake in de voorbeelden (5), maar minder in de zinnen in (6). In die gevallen is de matrix-structuur (de hoofdzin, of een substantief) *conceptueel afhankelijk* van de informatie in de bijzin. Een uitvoerige motivatie voor deze wijze van scores is te vinden in Schilperoord (1996, met name hoofdstuk 4), en in Schilperoord & Verhagen (1997b). De notie 'conceptuele afhankelijkheid' wordt daar als volgt gedefinieerd: 'if a constituent of clause A is conceptually dependent on a clause B, B is an integral part of the conceptualization of A, and therefore not available as a separate discourse segment, i.e. cannot enter into a discourse coherence relation with A, or any other part of the discourse'. Zo kan de referent van het substantief 'telefax' niet adequaat bepaald worden zonder de informatie in de relatieve zin. En in (6) heeft de adverbiale groep *in zoverre* tezamen met *verschillend* het effect dat een schaal gesuggereerd wordt met als uitersten 'totaal verschillend' en 'totaal niet verschillend'. De betekenis van 'in zoverre' kan echter niet bepaald worden zonder de 'kalibrerende' graad-clause

dat in *PLAATSNAAM*.... En omdat eenheden van planning (en productie) toch allereerst eenheden van conceptualisatie van uit te drukken inhouden lijken te zijn, is ervoor gekozen in gevallen als (6) geen clause-grens aan te nemen (hoewel die daar op syntactische gronden dus wel aangenomen kan worden). Optredende pauzes op die locaties werden daarom gescoord als 'tussen constituenten'.

Met betrekking tot de resultaten van de frequentie- en lengte-analyse is het van belang op te merken dat een directe vergelijking met de studie van Ford belemmerd wordt door het feit dat Ford's clause-begrip afwijkt van het hier gehanteerde. Haar analyse richt zich op de verschillen in proportie en lengte van pauzes op verschillende grammaticale locaties: binnen clauses, daarbinnen, en op zinsgrenzen. Ford's corpus betreft een reeks spontane monologen van 20 proefpersonen, handelend over onderwerpen als 'vrije tijdsbesteding', 'familie' en 'houding ten aanzien van het universitaire leven'. Bij het scoren van pauzeerlocaties maakte Ford onderscheid tussen wat zij 'basic clauses' en 'full finite clauses' noemt. De notie 'basic clause' refereert aan een 'sentoid' die in de oppervlakte-structuur al dan niet opduikt in de vorm van een structuur die wordt gedomineerd door een finiete werkwoordsvorm. Bezie een 'full finite clause' als (7)

(7) Jan vroeg Piet te gaan

Zin (7) bevat in Ford's optiek twee basic clauses: *Jan vroeg Piet* en *0 te gaan*. Daarbij is de tweede echter *niet* als een finiete clause in de oppervlakte-structuur gerepresenteerd. Het dispuut dat Ford wenste te beslechten betrof de vraag of het nu de full finite clause of de basic clause was die de beste manifestatie van de onderliggende productie-eenheid is. Men zou immers kunnen betogen dat de onderliggende representatie van '0 te gaan' de volledig opgetuigde clause-structuur 'Piet gaat' is, en dat die daarom ook als eenheid word gepland. De regel die er voor zorgt dat we *Piet* begrijpen als het subject van 'te gaan' wordt wel *object-raising* genoemd. Overigen zal duidelijk zijn dat de notie 'basic clause' zich slecht verdraagt met die van 'conceptuele afhankelijkheid'. In de gegeven voorbeeldzin is van conceptuele afhankelijkheid sprake omdat een betekenisvolle conceptualisering van het werkwoord 'vragen' niet mogelijk is zonder referentie aan het 'object' van vragen. Vandaar dat volgens het door mij gehanteerde clause-begrip aan de betreffende zin één conceptuele eenheid ten grondslag ligt. De ruimte staat het niet toe om op deze kwestie verder in te gaan, maar zie Schilperoord & Verhagen (1997b). Met inachtname van deze verschillen in locatie-scores in beide studies (die vanaf nu genegeerd zullen worden), kunnen we nu een blik werpen op de resultaten. Die staan samengevat in tabel 1.

Ford's data zijn getoetst met behulp van ANOVA (Ford 1982, 806). Significante verschillen betroffen die tussen de gemiddeldes voor zinspauzes, pauzes voor clauses en pauzes binnen clauses. Ook de proportie-verschillen bleken significant. Zij vond, zoals was gehypothetiseerd,

Tabel 1. Proportie en gemiddelde lengte van pauzes in mondelinge (F(ord)) en schriftelijke (S(chilperoord)) taalproductie voor de locaties 'tussen zinnen', 'tussen (basic) clauses' en 'binnen clauses'.

locaties	Gem. pauze lengte (ms)		Pauze proporties	
	F	S	F	S
zinnen	1601	2676	.659	.95
finiete clauses	962	1336	.235	.68
basic clauses	1034	--	.206	--
binnen clauses	872	1185 (con)	.095	.60 (con)
		1046 (woo)		.53 (woo)

geen verschil in lengte en proportie van pauzes tussen finiete en basic clauses. Mijn data zijn getoetst met behulp van Multi-level variantie-analyse (Schilperoord 1996: 117ev). De locaties verschilden zowel naar pauzeerlengte als -proportie significant. Er bestaat dus voor beide productiemodi een effect van locatie op de lengte en proporties van pauzes. Tabel 1 wijst verder uit dat voor beide productiemodi het algemene data-patroon nagenoeg gelijk is. Zowel de gemiddelde lengte als de proportie pauzes is voor beide productiemodi bij zinnen het hoogst: respectievelijk 1601ms en .659 voor mondelinge productie en 2677ms en .95 voor schriftelijke. Zinsgrenzen zijn derhalve de voornaamste plaatsen voor pauze-activiteiten. De binnen-clause locaties scoren in de twee modi op beide parameters het laagst: .872ms en .095 voor mondelinge productie en 1185/1046ms en .60/.53 voor schriftelijke. Op deze locatietypen vindt derhalve veel minder pauze-activiteit plaats. Clauses nemen voor beide modi en voor beide parameters een tussenpositie in: 962/1034ms en .235/.206 voor mondelinge productie en 1336ms en .68 voor schriftelijke.

Men zou nu, althans met betrekking tot schriftelijke tekstproductie, kunnen stellen dat hiermee het eenheidspleit is beslecht in het voordeel van de zin. Immers, woord- en constituentengrenzen, maar ook clause-grenzen, trekken in veel mindere mate pauze-activiteiten aan, en bovendien duren pauzes op die plaatsen minder lang. Zo'n conclusie is echter wat haastig zoals ik in het vervolg zal laten zien, en moet een voorbehoud bij de hier gerapporteerde data gemaakt worden (althans, bij mijn eigen data). Dat heeft betrekking op de *variantie-data* voor de pauzeerlengtes op de geanalyseerde locaties. De geobserveerde pauzes op die locaties wijken (uiteraard) af van de gemiddelden die daarvoor berekend zijn. De vraag is dan in welke mate daarvan sprake is, en hoe die varianties verklaard kunnen worden.

De variantie-analyse die op de data werd uitgevoerd wijst uit dat, ten eerste, de variantie in pauzeerlengte op zinsgrenzen veel groter is dan op clause-grenzen, en, ten tweede, dat die schommelingen op zinsgrenzen veel krachtiger samenhangen met verschillen tussen teksten, en tussen de alinea's waarin die zinnen optreden dan de schommelingen op clause-grenzen. Zulke verschillen wijzen dus uit dat, ten eerste, de lengte van pauzes tussen zinnen veel minder adequaat te voorspellen is op basis van dat locatie-type en, ten tweede, dat de clause als (mogelijke) eenheid

van productie veel minder vatbaar is dan zinnen voor (inhoudelijke) verschillen tussen, hetzij teksten, hetzij alinea's (zie verder Schilperoord 1996, met name hoofdstuk 3).

De 'eenheids' kwestie wordt nu een zaak van de interpretatie die men geeft aan de gemiddelde pauzelengte en variantie daarin. Kijk je (alleen) naar lengte en proportie, dan is de (vol-)zin de voor de hand liggende eenheidskandidaat. De variantie-data leiden echter tot een andere conclusie. Betoogd kan worden dat de eenheid van productie bepaald wordt door de structuur die pauze-activiteiten genereert die relatief invariant zijn ten opzichte van verschillen tussen teksten en tussen alinea's binnen teksten. De (vol-)zin is dan een minder voor de hand liggende kandidaat. Om meer licht op deze zaak te werpen, is gekozen voor een (naar mijn weten) nieuw type pauze-analyse; één waarin niet zozeer centraal staat wat er *op de grenzen* van de gehypothetiseerde eenheid aan pauze-activiteit plaatsvindt, maar juist *binnen* die eenheid. Zo'n analyse wordt beschreven in paragraaf 3.

3. Temporele modificatie in zinnen en clauses

3.1 Inleiding

Centraal in deze paragraaf staat de door Dwight Bolinger ingevoerde notie van *linear modification*, die door hem als volgt wordt omschreven:

'Before a speaker begins, the possibilities of what he will communicate are practically infinite, or, if his utterances are bound within a discourse, they are at least enormously large. When the first word appears, the possibilities are vastly reduced.' (Bolinger 1965, 219)

Bolinger geeft hier uitdrukking aan de gedachte dat ieder volgend woord in een uiting een reductie impliceert van de mogelijkheden die de spreker voor het vervolg ervan nog open staan. Stel dat we deze gedachte projecteren op een hypothetische tijdlijn waarop behalve de consecutieve reeks woorden ook pauzes tussen die woorden zijn weergegeven. We kunnen dan stellen dat de situatie waarin de mogelijkheden 'enormously large' zijn gekenmerkt is door een relatief hoge mate van 'onzekerheid' die echter, naarmate de spreker verder komt in de uiting, geleidelijk zal afnemen. Dit idee van geleidelijke onzekerheidsreductie kunnen we vertalen naar (predicties aangaande) de ontwikkeling van pauze-tijden. Met de afname in onzekerheid over continuering valt te verwachten dat pauzes ook geleidelijk *kort* gaan duren. Bovendien kunnen we aannemen dat de geleidelijk verhoogde mate van voorspelbaarheid ertoe zal leiden dat de pauzertijd-variantie geleidelijk zal afnemen (dus de lengte van een geobserveerde pauze kan steeds beter op basis van het algemeen gemiddelde voorspeld worden).

Deze predictie staat of valt echter met het *eenheidskarakter* van de structuur waarin pauzes optreden. Stel dat er sprake is van twee gehypothetiseerde eenheden x en y die corresponderen met structuren X en Y waarvoor geldt dat X bestaat uit één of meerdere Y 's, als in (8).

(8) $(\dots)_X(\dots)_Y\dots_X$

Als X de structuur is die correspondeert met de 'werkelijke' productie-eenheid, dan zal temporele modificatie (zo daar überhaupt sprake van is) optreden binnen X , onafhankelijk van eventuele Y -grenzen. Is daarentegen Y de structuur die correspondeert met de 'werkelijke' productie-eenheid, dan zal er uitsluitend, en wel van *herhaalde* temporele modificatie sprake zijn *binnen* all Y -structuren. Echter, op de Y -grenzen zullen de mogelijkheden voor het vervolg weer 'enormously large' zijn. We hypothetiseren derhalve dat *temporele modificatie* in de lengte van pauzes alleen optreedt binnen een structuur die correspondeert met een productie-eenheid. De hypothetische structuren waarvoor de hypothese hier onderzocht wordt zijn die van (vol-)zinnen en clauses. Eenheden *kleiner* dan de clause zouden uiteraard ook in zo'n analyse betrokken kunnen worden. Dat is echter nagelaten omdat we op zoek zijn naar de *grootst* mogelijke structuur die (nog) als productie-eenheid aan te merken valt.

3.2 Data-analysemodel

Om de temporele modificatie-hypothese te onderzoeken moest de variabele 'pauze-lengte' gemodelleerd worden als een tijdfunctie $f(t)$, met pauzertijd P als afhankelijke variabele en 'tijd' als verklarende variabele, zodat $P = f(t)$. De variabele 'tijd' werd gedefinieerd met als eenheden 'posities tussen woorden', en de analyse was gericht op de ontwikkeling van de parameters 'pauze-lengte' en 'pauze-variantie' voor elke opvolgende tussen-woord positie binnen clauses en zinnen. Het begrip 'temporele modificatie' is als volgt te operationaliseren:

- Van temporele modificatie is sprake bij een significante afname/toename van gemiddelde pauze-lengte voor de tussen-woord positie Y_w , ten opzichte van Y_{w-n} met $n \geq 1$ en $w = n + 1$.
- Van temporele modificatie is sprake bij een significante afname/toename van gemiddelde pauze-variantie voor de tussen-woord positie Y_w , ten opzichte van Y_{w-n} met $n \geq 1$ en $w = n + 1$.

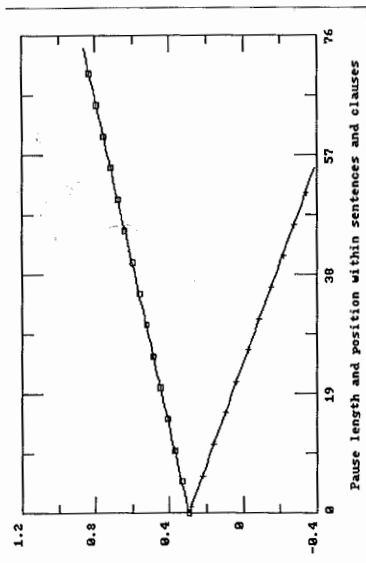
Alleen in geval van een significante *afname* kan volgens de temporele modificatie-hypothese worden aangenomen dat we met een *eenheid* van productie van doen hebben. Dat kan *niet* aangenomen worden in geval het effect afwezig is, of wanneer het tegenovergestelde effect optreedt (i.e. de pauzes worden gaandeweg langer en/of de variantie in pauze-lengte neemt gaandeweg toe).

3.3 Resultaten

In figuur 1 zijn de resultaten weergegeven voor de temporele ontwikkeling van gemiddelde pauzelengte voor zinnen (blokken) en clauses (kruisen). Op de x-as zijn de mogelijke pauzeertlocaties weergegeven (alle tussen-woord posities), op de y-as de gemiddelde lengte. Van belang is hierbij alleen het intercept, dat wil zeggen de gemiddelde lengte van de pauzes na het eerste woord in de zin/clause. Dit intercept werd vastgesteld op .65.

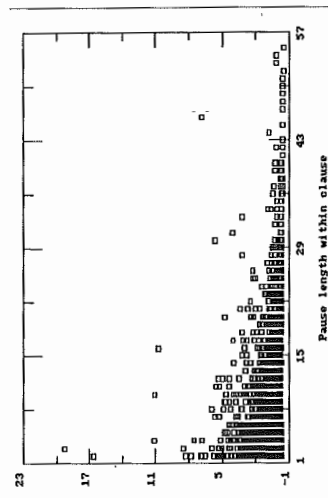
De regressielijnen voor zinnen en clauses snijden de verticale as op het intercept. De grafiek maakt zonneklaar dat er sprake is van een *graduele afname* in gemiddelde lengte voor elk opvolgende pauze binnen clauses. Merk op dat dit *niet* betekent dat de tweede pauze altijd *kort* (in clauses) of *langer* (in zinnen) is dan de eerste. Het betreft hier gemiddelde pauzelengtes voor pauzes tussen elk opvolgend woordpaar. De geobserveerde afname binnen clause-structuren is

Figuur 1. *Ontwikkeling van de gemiddelde pauzeerlengte naar lineaire positie in zinnen (blokken) en clauses (kruisjes). Er zijn zoveel mogelijke posities als woorden in de zin/clause.*



significant, dus de regressielijn voor pauzes binnen clauses 'daalt' reëel. De lengte van een pauze P op woordgrens n kan daarbij geschat worden als $P_{n-clause} = .65 - n * .013$, met .65 als het intercept en $n * .013$ de verminderde gemiddelde lengte voor de n-de pauze. Ook in zinnen is lineaire modificatie aanwezig, zij het exact in de tegenovergestelde richting. De toename is wederom significant, en de n-de pauze P in een zin heeft een gemiddelde lengte $P_{n-zin} = .65 + n * .008$. Deze resultaten vormen evidentie voor de gedachte dat inderdaad de clause, en niet de zin beschouwd moet als de (primaire) eenheid bij schriftelijke taalproductie. Extra steun hiervoor kan verkregen worden wanneer ook de ontwikkeling in variantie in pauzetijden als functie van tijd wordt geanalyseerd. De resultaten van deze analyse zijn weergegeven in figuur 2.

Figuur 2. *Ontwikkeling van variantie in pauzeerlengte binnen clauses.*

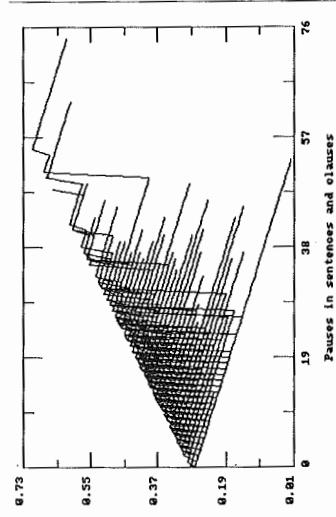


In de grafiek in figuur 2 representeert elk vierkant hokje een individuele pauze na het eerste woord, het tweede, enzovoorts binnen clauses (weer weergegeven op de x-as). Op de y-as is de lengte ervan in tienden van seconden weergegeven. De zwarte 'balkjes' in de figuur zijn het resultaat van een ophoping van blokjes, en hebben daarom geen betekenis. De grafiek onderstreep

de conclusie die op basis van de gemiddelde pauzeertijden werd getrokken: niet alleen worden pauzes in clauses gradueel korter; ook de varianties nemen af naarmate een pauze later in een clause optreedt. Met andere woorden, op basis van de schatting van de gemiddelde lengte kan gaandeweg steeds adequater de werkelijke lengte van een pauze voorspeld worden. De onzekerheidsfluctuaties doven als het ware uit naarmate de tekstproducent verder vordert binnen de clause. De afname in variantie bedraagt $n * .04$, en is wederom significant. De niet in een grafiek weergegeven toename in pauzeerlengte in zinnen is eveneens significant en bedraagt $n * .016$.

Tenslotte is getracht uit te zoeken hoe nu de simultane tendensen van *afname in clauses* en *toename in zinnen* voor beide parameters (lengte en variantie) verklaard kan worden. Deze vraag is niet triviaal: zinnen bestaan uiteindelijk uit clauses. De resultaten van deze analyse zijn weergegeven in figuur 3. Het intercept is wederom .65.

Figuur 3. *Pauzeertijd-ontwikkeling in zinnen en clauses.*



Figuur 3 laat twee opmerkelijke zaken zien. Allereerst blijkt dat het effect van graduele toename in gemiddelde pauze-lengte binnen zinnen geheel kan worden toegeschreven aan het feit dat pauzes tussen clauses steeds langer gaan duren naarmate de zin uit meerdere clauses is samengesteld. Hierop wijzen de verticale lijnen in de figuur. Verder blijkt dat temporele modificatie in clauses optreedt ongeacht de lineaire positie van die clause binnen de zin. Dit gegeven vormt naar mijn oordeel nog eens een krachtige aanwijzing dat clauses inderdaad de primaire productie-eenheid zijn. De temporeel bepaalde toename in pauzeduur tussen clauses kan wellicht verklaard worden uit het feit dat multi-clausale zinnen *complexer* worden naarmate die meer clauses omvatten. Gecombineerd lijken de resultaten op aardige wijze het principe van *rechtsuitbreidbaarheid* van Nederlandse zinsconstructies te illustreren. Kort gezegd, aan clauses komt verplicht een eind; aan zinnen in beginsel niet, maar het aanhechten van een nieuwe clause aan het al bestaande clause-complex gaat met steeds meer inspanning gepaard.

4. Discussie

De belangrijkste conclusie in deze bijdrage is dat de analyses steun bieden aan de gedachte dat clauses de primaire eenheid van schriftelijke tekstproductie zijn. De temporele modificatie-data

impliceren dat de proces-inspanning, zoals gesignaleerd door pauzes, gaandeweg de clause afneemt (figuur 1, paragraaf 3), om vervolgens, op de grens tussen twee clauses, weer sterk toe te nemen (figuur 3, paragraaf 3). Op die plaats neemt de gemiddelde pauze-lengte, alsmede de lengte-variantie aanzienlijk toe. Deze procesgang vertoont een 'zaagtand'-achtig verloop van graduele afname en abrupte toename in proces-inspanning. Teksproducenten plannen, zo lijkt het dus, hun teksten on-line in de vorm van conceptuele incrementen die corresponderen met clauses in de uiteindelijke tekst. Dit betekent ook dat zinnen, hoewel die op grond van parameters als pauze-lengte en frequentie duidelijk onderscheidbaar zijn als eenheden in de productie van teksten, in wezen op een clause-voor-clause wijze tot stand komen. Nader onderzoek zal moeten uitwijzen of, en zo ja in hoeverre, de geobserveerde varianties in pauzetijden binnen clauses verklaard kunnen worden op grond van de distinctie tussen hoofdzin (onafhankelijke clause) en bijzin (afhankelijke clause), en/of op grond van die tussen verschillende typen bijzinnen. Bovendien moet nagegaan worden in hoeverre Ford's notie van 'basic clauses' een betekenisvolle rol speelt in schriftelijke tekstproductie.

Literatuur

- Boomer, D.S. (1965). Hesitation and grammatical encoding. *Language and Speech*, 8, 148-158.
- Bolinger, D. (1965). Linear modification. In: Bolinger, D., *Forms of English; accent, morpheme, order*. Cambridge MA: MIT Press.
- Chafe, W. (1987). Cognitive constraints on information flow. In: Tomlin, R. (ed.), *Coherence and grounding in discourse*. Amsterdam: Benjamins.
- Ford, M. (1982). Sentence planning units: implications for the speaker's representation of meaningful relations underlying sentences. In: Bresnan, J. (ed.), *The mental representation of grammatical relations*. Cambridge MA: MIT Press.
- Ford, M. & V.M. Holmes (1978). Planning units and syntax in sentence production. *Cognition*, 6, 35-53.
- Holmes, V.M. (1988). Hesitations and sentence planning. *Language and Cognitive Processes*, 3, 323-361.
- Levelt, W.J.M. (1989). *Speaking: from intention to articulation*. Cambridge MA: MIT Press.
- Schilperoord, J. (1996). *It's about time; temporal aspects of cognitive processes in text production*. Amsterdam/Atlanta: Rodopi.
- Schilperoord, J. (1997). *Eenheden van tekstproductie: de clause als domein van temporele modificatie*. Lezing gehouden bij de oprichting van het Utrechts instituut voor Linguïstiek, UIL OTS, 13-2-1997
- Schilperoord, J. & Verhagen, A. (1997a). *Functionele elementen in een cognitief perspectief: evidentie uit taalproductie*. Te verschijnen in *Taalkunde*.
- Schilperoord, J. & Verhagen, A. (1997b). *Conceptual dependency and the clausal structure of discourse*. Te verschijnen in Koenig, J. (ed.), *Proceedings of the second conference on Conceptual Structure, Discourse and Language*. Chicago Press.

COGNITIEVE ACTIVITEITEN TIJDENS HET SCHRIJFPROCES EN RELATIES MET TEKSTKWALITEIT:

PLANNEN EN FORMULEREN

Gert Rijlaarsdam & Huub van den Bergh

Samenvatting

In veel studies wordt het schrijfproces van schrijvers gekarakteriseerd door middel van de frequentie waarmee een schrijver bepaalde cognitieve activiteiten onderneemt. Hiermee wordt impliciet aangenomen dat het moment waarop een activiteit uitgevoerd wordt niet relevant is. In deze bijdrage laten wij zien dat deze aanname niet terecht is; het moment in het schrijfproces waarop een schrijver een cognitieve activiteit onderneemt is van cruciaal belang. Dit geldt niet alleen om een adequate beschrijving te geven van het schrijfproces, of een onderscheid te maken tussen het schrijfproces van verschillende schrijvers, maar ook om relaties tussen het schrijfproces en de kwaliteit van de resulterende tekst aan te kunnen tonen.

1. Inleiding

In het begin van de tachtiger jaren richtten verschillende cognitieve psychologen zich op het schrijfproces als onderwerp van onderzoek (Flower & Hayes, 1980; Bereiter & Scardamalia, 1980). Vanaf die tijd is er als het ware een 'boom' in studies naar cognitieve processen of cognitieve activiteiten gedurende het schrijfproces, in plaats van onderzoek naar effectief (schrijf)onderwijs en de inrichting van de leeromgeving. Cognitieve psychologen en cognitief georiënteerde tekstproductie-onderzoekers in de VS (zie: Gregg & Sternberg (1980) of meer recentelijk Levy & Ransdell, 1996) en Europa (zie: Eigler & Jechle, 1994, of meer recentelijk Rijlaarsdam, Van den Bergh & Couzijn, 1996) hanteerden meestal het zogenaamde 'expert-novice' paradigma. In veel van deze studies zijn verschillen aangetoond tussen het schrijfgedrag van novices en experts (zie Rumes (1985) voor een overzicht). Het model van Flower & Hayes (1980) werd daarbij veelvuldig gebruikt als 'theorie' of interpretatiekader.

Eén van de meest bestudeerde subprocessen is 'revisie' (zie: Fitzgerald, 1987). De resultaten van de verschillende onderzoeken bleken helaas tegenstrijdig, iets dat indicatief is voor het hele onderzoeksgebied. Reviseren experts vaker dan novices, met name wanneer het revisies op globaal tekstueel niveau betreft, zoals sommige onderzoekers rapporteren (Faigley, 1981; Flower, Hayes, Carey, Shriver & Stratman, 1986; Monahan, 1982; Wright, 1983)? Of, reviseren experts minder vaak omdat zij beter plannen en/of beter reviseren (één goede revisie is beter dan vier matige revisies)?

Eén van de problemen van het schrijfonderzoek is het ontbreken van een notie van context. Hierbij kan aan het begrip 'context' twee interpretaties gegeven worden. De eerste interpretatie heeft betrekking op de experts in het expert-novice paradigma. Experts zijn experts op één gebied en op dit gebied zouden zij ander ('beter') schrijfgedrag ten toon kunnen spreiden dan novices. Op andere gebieden kunnen zij wellicht niet zonder meer vertrouwen op hun 'gewone' expertgedrag, omdat het genre of het onderwerp relatief onbekend zijn (zie Torrance,