

- Mann, W., & S. Thompson
1988 Rhetorical Structure Theory: toward a functional theory of text organization. In: *Text* 8: 243-281.
- Noordman, L.
1987 *Tekst in samenhang*. Openbare rede. Tilburg: KUB.
- Polanyi, L., & R. Scha
1983 The syntax of discourse. In: *Text* 3: 261-270.
- Redeker, G.
1990 Ideational and pragmatic markers of discourse structure. In: *Journal of Pragmatics* 14: 305-319.
- Sanders, T.
1990 Tekststructuur als criterium voor tekstkwaliteit. In: C. van Rees en P. Mooren (red.) *Makkelijk lezen*. Tilburg.
- Sanders, T.
1991 Tekststructuur in teksten van onervaren schrijvers: van tekstanalyse naar een criterium voor tekstkwaliteit. Te verschijnen in: *Toegepaste Taalwetenschap in Artikelen*.
- Sanders, T., & L. Noordman
1988 Tekststructuur en Begrijpelijkheid. *Begrijpelijkheidscriteria van ervaren tekstbeoordelaars*. In: *Tijdschrift voor Taalbeheersing* 10: 81-92.
- Sanders, T., W. Spooren & L. Noordman
1991 Towards a taxonomy of coherence relations. In: *Discourse processes*.
- Wijk, C. van
1991 Information analysis in written discourse. In: J. de Jong & L. Verhoeven (red.). *The construct of language proficiency*. Amsterdam.
- Wijk, C. van, & T. Sanders
1987 Linguistische criteria voor tekstkwaliteit. In: *Levende Talen* 426: 691-697.

8 Informatie in de tekst: een classificatie op basis van beschikbaarheid

Els van der Pool & Carel van Wijk

Inleiding

Naast structuur, vorm en verzorging is informatie een belangrijk aspect van een tekst. Het ligt daarom voor de hand bij het analyseren van een tekst niet alleen de structuur, vorm of verzorging in ogenschouw te nemen, maar ook de informatie aan een analyse te onderwerpen. Daarvoor zijn verschillende mogelijkheden.

In ons onderzoek bekijken we het aspect van informatie-verdeling in een tekst. Daarbij staat de vraag centraal in hoeverre informatie-verdeling de kwaliteit van een tekst bepaalt en in hoeverre informatie-verdeling beïnvloed wordt door schrijfvaardigheid.

Er zijn reeds een aantal methoden voor onderzoek naar informatie-verdeling. Ieder van die benaderingen heeft echter wel een aantal nadelen waardoor bij analyse van een tekst bepaalde zaken niet aan het licht komen. Door een aantal onderdelen uit bestaande methoden met elkaar te combineren en deze toe te passen op natuurlijk geproduceerde tekst is een nieuwe aanpak gecreëerd waarmee een diepgaande analyse mogelijk is, terwijl de originele tekst zo veel mogelijk intact blijft. Deze methode zal worden besproken.

Eerst worden de analyse-methode en de bijbehorende begrippen kort uiteengezet, daarna wordt de methode toegepast op een voorbeeldtekst. De voorbeeldtekst die gebruikt wordt, is onderdeel van een corpus van ongeveer 200 opstellen die dienen als onderzoeksmateriaal. De teksten zijn geschreven door 10-, 12- en 15-jarigen. De opdracht was: 'Beschrijf een persoon op wie je graag zou willen lijken'. Tot slot wordt kort aandacht besteed aan wat een dergelijke analyse aan gegevens over de tekst en over de schrijver kan opleveren.

1. Uiteenzetting analysemethode

Tot nu toe is bij informatie-analyse vooral gekeken naar de *eenheden* waarin de informatie is in te delen en welke functies die eenheden hebben (denk bijvoorbeeld aan de propositie-analyse van Kintsch & Van Dijk (1978) en de case grammar van Fillmore (1968)). Tevens is bekeken in hoeverre de informatie *nieuw* is ('given'-'new' van Halliday (1967), Chafe (1987) en Prince (1981)).

In onze aanpak worden beide benaderingen geïntegreerd (zie ook Van Wijk 1991): eerst wordt de tekst opgesplitst in informatie-eenheden, concepten en bepalingen, waarna de eenheden worden geclassificeerd naar de mate waarin ze nieuw of te wel beschikbaar zijn.

De achterliggende gedachte bij de analyse-methode is dat uitingen (zinnen of zinsdelen, clauses) gegroepeerd zijn rond een propositionele kern die bestaat uit concepten. Met een concept bedoelen we een 'idee-eenheid' m.b.t. een object, actie, gebeurtenis, proces of eigenschap. Neem bijvoorbeeld (1):

- (1) Tom drinkt altijd dure wijn want hij houdt niet van goedkope.

Deze zin bestaat uit twee deelzinnen: 'Tom drinkt altijd dure wijn' en 'hij houdt niet van goedkope'. De kern van het eerste conjunct is 'Tom drinkt wijn'. Deze propositionele kern bestaat uit drie concepten 'drinken' 'Tom'-'wijn'. De kern van het tweede conjunct is 'Tom houdt van wijn' en bestaat eveneens uit drie concepten 'houden' van-'Tom'-'wijn'.

Zoals blijkt uit de voorbeeldzin corresponderen concepten niet altijd 1 op 1 met woorden, want in plaats van 'Tom drinkt altijd dure wijn' had er ook kunnen staan 'Tom heeft vaak trek in wijn' waarbij 'trek-hebben-in' één concept is. Verder blijkt dat verwijzingen naar concepten niet noodzakelijk expliciet zijn. Er kan gebruik worden gemaakt van referentiële expressies, zo verwijst het pronomen 'hij' in de tweede deelzin naar het concept 'Tom' uit het eerste deel.

De propositionele kern van een zin kan worden uitgebreid door middel van bepalingen. Er zijn drie soorten bepalingen, gedefinieerd naar het bereik dat ze hebben. Allereerst zijn er bepalingen bij een concept zoals 'dure' bij 'wijn' in het eerste conjunct en 'goedkope' bij 'wijn' in het tweede conjunct van zin (1). Deze bepalingen worden vanwege hun bereik *conceptual modifiers* genoemd. Ten tweede zijn er bepalingen bij een *propositie* zoals 'altijd' bij de propositie 'drinken-Tom-wijn'. Deze bepalingen worden *qualifiers* genoemd: ze kwalificeren een propositie. Voorbeelden van dit type bepaling zijn tijds- en plaatsbepalingen en negaties.

Ten derde zijn er bepalingen die *twee proposities aan elkaar relateren*. Een voorbeeld hiervan is 'want' dat de twee deelzinnen uit (1) 'Tom drinkt altijd dure wijn' en 'hij houdt niet van goedkope wijn' met elkaar verbindt. Deze bepaling wordt een *relation modifier* genoemd. 'Want' geeft immers de relatie aan tussen de eerste en tweede propositie. Relation modifiers komen voor in de vorm van connectoren en sequentie-aanduiders.

Indeling in concepten en bepalingen is slechts de helft van het werk bij informatie-analyse. Wat ook bepaald moet worden is de mate waarin een concept nieuw ofwel beschikbaar is. Bij het bepalen van de mate van beschikbaarheid wordt bekeken in hoeverre de schrijver rekening heeft gehouden met de beschikbaarheid van concepten bij de lezer. Het standpunt dat wordt ingenomen bij de analyse is dus vanuit de schrijver (tekstproductie) en niet vanuit de lezer (verwerking).

De beschikbaarheid wordt uitgedrukt in de mate waarin een concept al dan niet actief is. Er zijn drie maten van activatie die zich op een continuum bevinden, resp. actief, semi-actief en inactief.

Een concept is *actief* als het reeds geïntroduceerd is in de tekst. Het vormt de beschikbare, 'oude' informatie.

Zowel reeds geïntroduceerde als nog niet-geïntroduceerde concepten kunnen *semi-*

actief worden c.q. zijn. Een reeds geïntroduceerd concept is semi-actief als het 'te lang geleden' genoemd is om nog actief te zijn. Het concept is dan gedeactiveerd. Een nog niet-geïntroduceerd concept is semi-actief als het beschikbaar is op basis van kennis van de wereld (bijv. georganiseerd in schema's of frames), of semantische velden (bijv. weergegeven in deel-geheel-relaties, hypernym-hyponym).

Om te bepalen wat dat kennisdomein precies omvat, is voor het gebruikte corpus een lijst samengesteld met kenmerken van personen zoals o.a. sexe, leeftijd, uiterlijk, werk, karakter, etc.. Op basis van die kenmerken wordt bepaald of een nog niet-geïntroduceerd concept al dan niet afleidbaar c.q. semi-actief is. Tot het kennisdomein behoort ook de 'kennis' die we hebben over de schrijver en de 'kennis' uit de instructietekst.

Een concept is *inactief* als het nog niet eerder genoemd is in de tekst of als het niet afleidbaar is. Het vormt de nieuwe informatie.

De mate waarin een concept actief of beschikbaar is kan veranderen per (deel-)zin. Neem bijvoorbeeld het concept 'Tom' uit (1).

- (1) Tom drinkt altijd dure wijn want hij houdt niet van goedkope.

In het eerste conjunct van (1) worden 'Tom' en 'wijn' geïntroduceerd en zijn inactief. In het tweede conjunct zijn ze niet langer inactief maar actief. Dit heeft tot gevolg dat 'Tom' gepronominaliseerd kan worden en 'wijn' door middel van een samen-trekking kan worden gedeleerd.

2. Toepassing van de analyse-methode

De voorbeeldtekst, geschreven door een 10-jarige jongen, is afkomstig uit het eerder genoemde corpus. Hij heeft de persoon beschreven op wie hij graag wilde lijken, meneer Ooms. Hieronder wordt de tekst gegeven in de segmentering waar in de analyse van wordt uitgegaan.

- | | |
|----|--|
| 1 | Ik zou graag willen lijken op Meneer Ooms. |
| 2 | Hij is van beroep politie-agent. |
| 3 | Ongeveer 40 jaar oud. |
| 4 | Als ontspanning vangt hij meestal vissen. |
| 5 | Maar ik zou liever voetballen. |
| 6 | Hij wordt niet heel gauw kwaad |
| 7 | maar |
| 7a | als hij kwaad is, |
| 7 | is hij het goed. |
| 8 | Hij is nogal groot |
| 8a | ongeveer 1.75 m. tot 1.80 m. |
| 9 | Hij houdt van vogels |
| 10 | en heeft zelf ook een hele grote volière. |
| 11 | Ook is hij bij de vogelvereniging. |
| 12 | Hij heeft er ook al prijzen mee gewonnen. |
| 13 | Ik zou graag op hem willen lijken vanwege zijn beroep en om zijn karakter. |

Bij de analyse worden de concepten en bepalingen in verschillende kolommen gezet. In de eerste kolom staan de relation-modifiers, in de tweede kolom de qualifiers. De conceptuele bepalingen staan in de kolommen bij de concepten (aangegeven d.m.v. '<'). Deze staan in kolom 3 t/m 5; in kolom 3 staan de actieve, in kolom 4 de semi-actieve en in kolom 5 de inactieve concepten.

In tabel 1 staat het resultaat van de tekst in zijn geanalyseerde vorm. Het voert te ver alle geanalyseerde zinnen te bespreken, we zullen ons daarom beperken tot de zinnen 1-2-3, 6-7a-7 en 13.

Tabel 1. De analyse van de voorbeeldtekst

Bepalingen		Concepten		
Relatie	Qualifier	Actief	Semi-actief	Inactief
1	T, zal, willen graag	Hij	Ik, lijken-op- ϕ	ϕ =Meneer Ooms
2			van-beroep-zijn- ϕ	ϕ =politie-agent
3			jaar-oud#	#=40<ongeveer
4	meestal	ϕ	als-ontspanning	vis-vangen
5	T, zal, willen graag	ik		voetballen
6	liever wordt, niet heel-gauw	hij		kwaad
7a	goed	hij, kwaad		
7		hij, het		
8		hij		
8a		ϕ	meter#	groot<n-ogal
9		hij		#=1.75-1.80
10	en, ook	ϕ , zelf		<ongeveer houden-van- ϕ
11	ook	hij		ϕ , vogels volière< hele- grote zijn-bij- ϕ
12	ook	hij, ϕ =er		ϕ , ϕ =vogel- vereniging prijzen- winnen-met- ϕ
13	vanwege, en, om	ik, ϕ =hem	lijken-op- ϕ	
		#1=zijn	beroep#1	
		#2=zijn	karakter#2	

De eerste zin, 'Ik zou graag willen lijken op meneer Ooms', bevat als concepten 'ik', 'lijken-op', 'meneer Ooms'. De concepten 'ik' en 'lijken-op' zijn afleidbaar uit de

instructie-tekst en kennis over de schrijver en zijn daardoor semi-actief. Werkwoordelijke uitdrukkingen (trek-hebben in uit (1)) en werkwoorden met een vast voorzetsel (lijken-op) eisen de aanwezigheid van een object, recipient, etc.. Dat wordt in de tabel aangegeven door middel van '\$'. Het verplichte object bij 'lijken-op' is ϕ =Meneer Ooms dat de nieuwe informatie in de zin weergeeft.

'Lijken-op' wordt gemodificeerd door de hulpwerkwoorden 'zou' en 'willen'. Omdat het tijdsaspect ook een modificerende werking heeft, worden temporele vervoegingen in de tabel aangegeven met een 'T' als ze afwijken van de o.t.t. (onderscheid enkelvoud/meervoud blijft hierbij gehandhaafd). Het bijwoord 'graag' specificceert de gehele uitdrukking 'ik zou willen lijken op...' en is daarom een qualifier.

In zin 2, 'Hij is van beroep politie-agent', verwijst 'hij' naar 'meneer Ooms'. Beide verwijzen naar hetzelfde concept maar de mate waarin het concept actief is, verandert in zin 2. Omdat 'hij' afleidbaar is uit de voorafgaande tekst is het concept niet langer inactief maar actief.

Omdat het al dan niet hebben van een beroep tot de domeinkennis behoort is 'van-beroep-zijn- ϕ ' semi-actief. Het heeft als object 'politie-agent' bij zich, wat nieuwe informatie is. 'Politie-agent' komt daarom in de kolom met inactieve concepten.

In de derde zin, 'Ongeveer 40 jaar oud', ontbreekt het onderwerp. In de tabel wordt dit door middel van een ' ϕ ' aangegeven. Het teken geeft aan dat er een element door samentrekking of ellipsis is komen te vervallen.

Informatie omtrent de leeftijd van meneer Ooms is bevat in het concept '40 jaar oud'. Het is één concept dat echter qua lexicale vorm opsplitsbaar is voor wat betreft de beschikbaarheid. Omdat leeftijd deel uitmaakt van de domeinkennis is 'jaar-oud' semi-actief. '40' is hierbij de nieuwe inactieve informatie. Het is de waarde die aan het kenmerk leeftijd wordt toegekend. Op deze manier is het mogelijk nieuwe informatie aan oude te koppelen. Deze koppeling wordt in de tabel weergegeven door een '#'. Met de toevoeging 'ongeveer' geeft de schrijver blijk van enige twijfel omtrent de precieze leeftijd. 'Ongeveer' is een bepaling bij '40-jaar-oud'. Dergelijke conceptuele bepalingen zijn in de tabel aangegeven door middel van '<'. Zin 4 en 5 laten we buiten deze bespreking omdat daar analytisch gezien weinig interessante aspecten in voorkomen.

In de zesde zin, 'Hij wordt niet heel gauw kwaad', is het concept meneer Ooms waar met 'hij' naar gerefererd wordt, nog steeds actief. 'Kwaad' vormt de nieuwe informatie en wordt gekwalificeerd door 'wordt' en 'heel gauw', hetgeen weer opsplitsbaar is maar achterwege kan worden gelaten. De gehele propositie wordt ontkend door de qualifier 'niet'.

Omdat bekend is vanuit zin 6 dat meneer Ooms kwaad kan zijn, is 'kwaad' in 7a van inactief actief geworden. De activatie van 'hij' verandert in 'als hij kwaad is' niet en blijft dus actief.

'Als' geeft in deze deelzin de voorwaarde aan waaraan moet zijn voldaan wil het in 7 genoemde gevolg optreden. Het verbindt op deze manier 7a aan 7 en vervult de functie van een relation modifier.

Via 'maar' in 'maar is hij het goed.' wordt 7 (en de in 7 ingebedde 7a) verbonden met 6. De relation modifier 'maar' geeft de tegenstellingsrelatie aan tussen beide zinnen. De activatie van het concept meneer Ooms, dat nog steeds wordt aangeduid met 'hij' is wederom niet veranderd. Naar het in 7a actieve 'kwaad' wordt nu gerefererd door middel van 'het', waarbij de activatie niet verandert. De mate waarin meneer Ooms kwaad is wordt aangegeven door de kwalificer 'goed'. Zin 8 t/m 12 laten we om dezelfde reden als zin 4 en 5 onbesproken.

Ten slotte zin 13: 'Ik zou graag op hem willen lijken vanwege zijn beroep en om zijn karakter.'

Gezien de voorafgaande zinnen zijn de concepten 'ik' en 'hem' allebei nog actief. 'Lijken-op' is, hoewel het in het voorafgaande (zin 1) genoemd is, hier niet actief. De reden daarvoor is dat na zin 1 enkele topic-switches hebben plaats gehad. Dat houdt in dat het onderwerp waarop de uitingen betrekking hebben niet constant is gebleven: in de eerste zin staat het 'lijken-op' centraal, daarna wordt er echter achtereenvolgens geschreven over het beroep van Meneer Ooms, zijn leeftijd, hobbies, karakter en weer over zijn hobby. Tot slot komt de schrijver terug op het 'lijken-op'. In zo'n geval is het 'te lang geleden' dat het concept genoemd is, en wordt verondersteld dat het niet actief is (gebleven) maar semi-actief is (geworden). Met andere woorden het concept is gedeactiveerd. 'Lijken-op' wordt evenals in zin 1 gemodificeerd door de hulpwerkwoorden 'zullen' en 'willen' en het bijwoord 'graag'.

De uiting tot zover behelst een bewering. Deze bewering wordt onderbouwd door twee argumenten die op gelijk niveau staan. Argument 1 luidt 'vanwege zijn beroep' en argument 2 luidt 'om zijn karakter'. We bespreken eerst de concepten in beide argumenten en daarna de bepalingen.

Argument 1: voor 'zijn-beroep' geldt gedeeltelijk hetzelfde als voor 'lijken-op'. De lexicale vorm is op te splitsen in 'zijn' en 'beroep'. 'Zijn' geeft de specifieke waarde aan van het kenmerk 'beroep' (analoog aan '40' bij 'jaar-oud') en slaat direct terug op 'hij' uit de vorige zin. 'Zijn' is daardoor i.t.t. 'beroep' wél actief. Dat 'beroep' semi-actief is wordt veroorzaakt door de topic-switches; het concept is inmiddels gedeactiveerd.

Argument 2: Het concept 'zijn karakter' is, net als 'zijn-beroep', opsplitsbaar. 'Karakter' is op basis van topic-switches semi-actief. Het wordt door middel van het reeds actieve 'zijn' verbonden aan de actieve informatie. Elk van de twee argumenten wordt met een eigen relation modifier aan de bewering verbonden. 'Vanwege' verbindt het eerste argument met de bewering en 'om' het tweede. 'En' functioneert als relation modifier die de beide argumenten onderling met elkaar verbindt.

3. Wat levert de analyse op?

Op de eerste plaats ontstaat na analyse van de tekst enig inzicht in de manier waarop naar concepten wordt verwezen. Neem bijvoorbeeld het concept 'Meneer Ooms' uit de voorbeeldtekst; de schrijver verwijst steeds naar hem door middel van 'hij', terwijl

er ook andere verwijzingen denkbaar zijn zoals 'die agent', 'deze man' of 'die meneer'.

Ten tweede kan na analyse de complexiteit van de concepten bekeken worden. Een concept kan complex zijn qua *opbouw* en qua *koppeling van informatie*.

Er is sprake van een complexe, gelaagde opbouw als een concept is geformeerd uit verschillende concepten. Een voorbeeld hiervan is het concept 'de heel erg oude politie-agent met de grijze haren en het blauwe pak'. Het is één concept dat is opgebouwd uit de concepten 'oude', 'politie-agent', 'grijze', 'haren', 'blauwe' en 'pak' die hiërarchisch geordend zijn. Een concept wordt complexer naarmate de hiërarchische structuur gelaagder is. In de voorbeeldtekst zijn nauwelijks voorbeelden te vinden van complex opgebouwde concepten.

Een andere vorm van complexiteit heeft betrekking op de koppeling van nieuwe aan oude informatie. Die koppeling kan vrij eenvoudig geschieden door een inactief concept met behulp van een werkwoord te verbinden aan actieve, oude informatie, zoals bijvoorbeeld in zin 8 gedaan wordt. Maar een iets gecompliceerdere vorm is ook mogelijk. Hiervan is sprake als nieuwe, inactieve informatie wordt geïntroduceerd op basis van semi-beschikbare informatie. Een voorbeeld daarvan is zin 3. Deze wijze van koppeling duidt op een ander informatie-management van de schrijver.

Deze twee zaken, verwijzing naar concepten en complexiteit van concepten, hebben gevolgen voor de begrijpelijkheid en aantrekkelijkheid van een tekst. Kiest een schrijver bijvoorbeeld voor eenduidige verwijzingen, waardoor de tekst begrijpelijker wordt, dan neemt de stilistische variatie af wat de aantrekkelijkheid van de tekst verkleint. Hetzelfde geldt voor de complexiteit van concepten.

Een derde punt dat wij hier tot slot willen noemen, is de spreiding van informatie over kolommen. Met betrekking tot concepten staat daarbij centraal of de tekst veelal nieuwe informatie bevat, zoals bij de voorbeeldtekst het geval is, of dat er ook een beroep wordt gedaan op bij de lezer aanwezige voorkennis door het gebruik van semi-actieve concepten.

Wat betreft de bepalingen kan worden bekeken in hoeverre de nieuwe informatie bestaat uit het leggen van verbanden tussen nieuwe en bekende informatie, dat wil zeggen in hoeverre de bepalingen in nieuwe informatie voorzien.

Literatuur

- Chafe, W.
1987 Cognitive constraints on information flow. In: R. Tomlin (Ed.), *Coherence and Grounding in Discourse*. Amsterdam. 21-51.
- Filmore, C.
1968 The case for case. In: E. Bach & R. Harms (Eds.) *Universals in Linguistic Theory*. London. 1-88.
- Halliday, M.
1967 Notes on transitivity and theme in English: Part 2. In: *Journal of Linguistics* 3, 199-244.
- Kintsch, W., & T. van Dijk
1978 Toward a model of text comprehension and production. In: *Psychological Review* 85, 363-394.
- Prince, E.
1981 Toward a taxonomy of given-new information. In: P. Cole (Ed.), *Radical Pragmatics*. New York. 223-255.
- Wijk, C. van
1991 Information analysis in written discourse. In: J. de Jong & L. Verhoeven (Eds.), *The Construct of Language Proficiency*. Amsterdam.