Van tekst naar prosodie

Hugo Quené en René Kager

Prosodische markeringen, zoals pauzering en accentuering, helpen luisteraars bij het begrijpen van spraak. In dit artikel bekijken we in hoeverre het mogelijk is om deze markeringen (voor pauzes en accenten) aan te brengen in teksten. Daartoe maken we gebruik van de prosodische zinsstructuur, die we proberen te achterhalen met minimale syntactische analyse. Vervolgens gaan we in op de fouten die dit algoritme nog maakt, en op de (on)mogelijkheden om die fouten op te lossen.

1 Inleiding

Gedurende vele eeuwen heeft de mensheid de bewegingen gevolgd van de 'dwaalsterren' of planeten, die zich ogenschijnlijk willekeurig bewegen door de nachtelijke sterrenhemel. Deze observaties hebben uiteindelijk geleid tot een toetsbaar model van ons zonnestelsel, zoals we dat nu kennen. Op basis van dat model kunnen we voorspellingen doen over de toekomstige positie van de planeet Mars in de sterrenhemel en vervolgens controleren (observeren) of deze planeet inderdaad op het juiste tijdstip op de juiste plaats aan de hemel te zien is. Onze voorspelling zal correct zijn, indien ons model rekening houdt met alle relevante factoren (de posities van de Aarde en Mars in hun respectievelijke banen om de zon, de rotatie van de Aarde, en de hoek van de aardas ten opzichte van zijn baanvlak).

Evenzo kunnen we een spraaksynthesesysteem (een 'sprekende computer') beschouwen als een model: niet van het zonnestelsel, maar van de menselijke spraakproduktie. Indien ons model alle relevante factoren omvat (en indien er geen technische beperkingen zijn bij de implementatie van ons model), dan moet de resulterende (synthetische) spraak niet te onderscheiden zijn van natuurlijke spraak. De observaties kloppen echter niet met de voorspelling: synthetische spraak is aanzienlijk slechter van kwaliteit dan natuurlijke menselijke spraak. We hebben blijkbaar onvoldoende rekening gehouden met bepaalde factoren die de menselijke spraakproduktie beïnvloeden. De planeet beschrijft weliswaar de voorspelde baan in de sterrenhemel, maar niet op het voorspelde tijdstip. Ons model verwaarloost een of meer belangrijke factoren.

Gevolg van deze tekortkomingen van ons model is dat de synthetische spraak slechter te verstaan is dan normale menselijke spraak en dat de boodschappen die de computer uitspreekt slechter begrepen en onthouden worden. Om een boodschap te kunnen begrijpen, moet de luisteraar de grammaticale zinsstructuur van de gesproken boodschap reconstrueren. De beide zinnen 1 en 2 beschrijven immers verschillende gebeurtenissen en – wat belangrijker is – dit verschil wordt uitsluitend aangegeven door de grammaticale verschillen tussen beide zinnen.

Hij geeft Frank een boek.
Frank geeft hem een boek.

In natuurlijke spraak komt deze grammaticale zinsstructuur mede tot uiting in accentuering van bepaalde woorden, intonatie, klinkerreductie, tempowisselingen, clitisatie, pauzering enz. (en daarnaast uiteraard in de woordvolgorde). Te zamen worden deze verschijnselen wel aangeduid met de term *prosodie*. Deze prosodische verschijnselen markeren de grammaticale zinsstructuur van de gesproken boodschap, en helpen de luisteraar bij het achterhalen van die structuur.

Bij synthetische spraak stelt het ontbreken van zulke markeringen van de zinsstructuur de luisteraar voor grote problemen (zie bijv. Collier en 't Hart (1975), Nooteboom en Kruyt (1987)). Het gaat hier blijkbaar om een belangrijke factor, die we in ons model niet zomaar mogen verwaarlozen. Synthetische spraak, de uitvoer van ons model, is niet acceptabel als deze markeringen ontbreken. We moeten dus rekening houden met de grammaticale zinsstructuur van de boodschap die 'door ons model gaat': de uitvoer dient markeringen te bevatten die deze zinsstructuur aangeven.

Een belangrijke vraag is natuurlijk, *welke* structuur we moeten weergeven in synthetische spraak. Welke zinsstructuur levert de beste voorspellingen op van prosodisch-fonetische verschijnselen als boven genoemd? Is dat de klassieke zinsontleding (in onderwerp, persoonsvorm, bijwoordelijke bepaling enz.) of hanteren sprekers en luisteraars een ander systeem om zinnen te analyseren? Deze vraag proberen taalkundigen de laatste jaren te beantwoorden. Het lijkt er sterk op dat niet zozeer de klassieke syntactische ontleding bepalend is voor prosodische markeringen, maar veeleer de (daarvan afgeleide) *prosodische zinsstructuur* (Nespor en Vogel (1983, 1986), Selkirk (1984), Kaisse (1985)). In par. 2 zullen we hierop nader ingaan.

Het is overigens onduidelijk of alle verschillende prosodische spraakkenmerken inderdaad gerelateerd zijn aan één enkele prosodische zinsstructuur. Wellicht heeft iedere zin een pauzestructuur, een intonatiestructuur en een tempostructuur, die onderling onafhankelijk zijn. Toch is dat niet zo aannemelijk. We gaan er voorlopig maar van uit dat alle prosodische verschijnselen in menselijke spraak bepaald worden door één enkele prosodische zinsstructuur. Experimentele evidentie voor deze aanname ontlenen we

Serie Spraak- en Taaltechnologie
Tekst-naar-Spraaksystemen

INFORMATIE JAARGANG 31 NR. 7/8
PAG. 559 T/M 636 1989

aan de observatie (Scharpff (1988)) dat discrepantie tussen twee prosodische spraakkenmerken (nl. intonatie of pauzering) een negatief effect heeft op de verstaanbaarheid van synthetische spraak.
Op grond van deze aanname is het dan gerechtvaardigd, om prosodische spraakkenmerken te interpreteren als markeringen van 'de' onderliggende prosodische zinsstructuur. Omgekeerd kan in synthetische spraak de afgeleide prosodische zinsstructuur toegepast worden bij het aanbrengen van prosodische kenmerken in de uitgevoerde spraak.

Ons model van de menselijke spraakproduktie, de sprekende computer, dient dus de correcte prosodische zinsstructuur te bepalen van de boodschap die uitgesproken moet worden. Op basis van die zinsstructuur kunnen dan vervolgens prosodische markeringen (pauzes, accentuering, tempowisseling enz.) aangebracht worden in de gesproken uitvoer. In dit artikel zullen we ons concentreren op het eerste probleem, het bepalen van de prosodische zinsstructuur, oftewel de prosodische zinsanalyse. In onze planeetanalogie kunnen we het probleem als volgt stellen: we weten dat de rotatie van de Aarde invloed heeft op de voorspelling van de positie van Mars aan de sterrenhemel. In eerste instantie beperken we het probleem: de rotatie van de Aarde dienen we exact te bepalen.
De prosodische zinsstructuur geeft ons een middel in handen, om de accentuering van woorden in een zin te bepalen (welke woorden moeten geaccentueerd worden). Dat is van belang, omdat correcte accentuering de verstaanbaarheid van synthetische spraak aanzienlijk verhoogt. Theoretisch worden accenten (voor een belangrijk deel) afgeleid uit de grammaticale zinsstructuur (Baart (1987); Gussenhoven (1984)). Deze theoretische benadering levert echter een aantal problemen op, waarop we hieronder in zullen gaan. In de praktijk kunnen we accenten ook wel afleiden uit de prosodische zinsstructuur, mits we genoegen nemen met wat meer accentfouten. We gaan daarom ook in op de 'toepassing' van de prosodische zinsstructuur bij het bepalen van de accentuering.

2 Prosodische zinsstructuren

De prosodische zinsstructuur van een zin is een hiërarchische representatie (boomstructuur), opgebouwd uit fonologische constituenten. Voor de bepaling van deze constituenten is het onderscheid tussen inhoudswoorden ('content words', hierna CW genoemd) en functiewoorden (hierna FW genoemd) van cruciaal belang. De groep inhoudswoorden wordt in theorie gevormd door nomina, werkwoorden, adjectiva en adverbia. Inhoudswoorden zijn de belangrijkste dragers van de inhoud of betekenis van een zin; uitbreidingen van de woordenschat van een taal betreffen ook voornamelijk inhoudswoorden (*raket, inschatten, jofel*). De groep functiewoorden wordt gevormd door de resterende syntactische woordsoorten (bijv. lidwoorden, voorzetsels, voegwoorden enz.). Deze woorden zijn niet essentieel voor de betekenis van een zin, maar fungeren vooral als het grammaticale 'bindmiddel' tussen de inhoudswoorden. Nieuwvormingen komen in deze categorieën nauwelijks voor.
De prosodische zinsstructuur is opgebouwd uit prosodische constituenten. Voor ons zijn twee van belang: de (fonologische) woordgroep (aangeduid als *Phi*) en de intonatiewoordgroep (*Int.*). Phi-constituenten komen ruwweg overeen met syntactische woordgroepen, hoewel er grote afwijkingen mogelijk zijn. De kern van zo'n Phi-constituent is het 'syntactische hoofd' (CW); andere 'toevoegingen' kunnen aan deze kern voorafgaan (zie 3). De algemene vorm van een Phi-constituent is dus als in 4.$[X]^n$ geeft aan dat X nul tot n keer kan voorkomen.

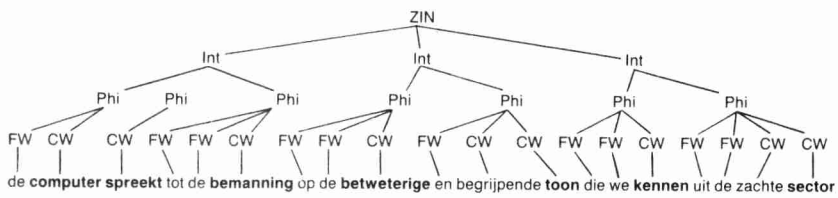
3 uit een heel dik *boek*

FW FW CW CW CW

4 $([FW]^n[CW]^nCW)$

De Phi-constituenten in een zin worden samengenomen onder één of meerdere Int-constituenten. Deze laatste constituent vormt het domein voor het intonatiecontour. Daarnaast is het een temporele organisatie-eenheid: zware grenzen (pauzes) liggen tussen Int-constituenten. De opdeling van een zin in deze constituenten wordt enerzijds bepaald door het syntactisch verband tussen de Phi-constuenten, maar anderzijds is het afhankelijk van sprekervariabelen als tempo en stijl. Figuur 1 toont tenslotte de volledige prosodische zinsstructuur van een Nederlandse zin. In dit voorbeeld wijzen we op de twee Phi-constituenten *op de betweterige* en *die we kennen,* die beide niet corresponderen met een syntactische woordgroep.

Volgens deze (gesimplificeerde) theorie wordt de prosodische zinsstructuur dus afgeleid uit de syntactische zinsstructuur en wel volgens vastgelegde regels (Nespor en Vogel (1983, 1986)). Deze werkwijze is ook denkbaar in ons model van de menselijke spraakproduktie, het spraaksynthesesysteem: eerst wordt een (uit te spreken) zin syntactisch ontleed. Vervolgens wordt uit deze ontleding een prosodische zinsstructuur afgeleid en tenslotte bepaalt deze prosodische zinsstructuur een groot aantal prosodische kenmerken in de resulterende spraak. Toch levert deze



Figuur 1: Prosodische zinstructuur en woordclassificatie van een Nederlandse zin; kernen van Phi-constituenten zijn vet gedrukt

werkwijze in de praktijk wel problemen op; deze betreffen vooral de syntactische ontleding van een zin. Het belangrijkste probleem wordt gevormd door de vele mogelijke ambiguïteiten in de grammaticale zinsstructuur. Toch moet die grammaticale zinsstructuur eerst bepaald worden, voordat we daaruit de prosodische zinsstructuur kunnen afleiden. Het gaat bij die ambiguïteiten doorgaans om de onderlinge syntactische verhoudingen tussen constituenten (zie Bakker (1989)). In voorbeelden 5 en 6 kan de verrekijker (of stok) een instrument van mij zijn ofwel een attribuut van de man.

5a (dat ik) (de man) ((met de verrekijker) (zie))
5b (dat ik) ((de man) (met de verrekijker)) (zie)

6a (dat ik) (de man) ((met de stok) (sla))
6b (dat ik) ((de man) (met de stok)) (sla)

7a (die) (het beleid) ((op de lange termijn) (uitwerkt))
7b (die) ((het beleid) (op de lange termijn)) (uitwerkt)

Beide versies van deze zinnen worden opgedeeld in identieke constituenten. Het verschil is gelegen in de syntactische verbanden tussen deze constituenten: Is *met de verrekijker* nu een bepaling bij (hoofdwerkwoord) *zie* of bij (lijdend voorwerp) *de man?* De oplossing van dit soort ambiguïteiten vereist een semantische en pragmatische analyse, waarbij het (niet-taalkundige) feit dat men kan zien met een verrekijker of kan slaan met een stok, een belangrijke rol speelt. Immers, de volgende zin is niet ambigu, omdat een stok geen instrument vormt bij het kijken (maar een verrekijker wel).

8 (dat ik) (de man) (met de stok) (zie)

Op dit moment is de vereiste semantische en pragmatische analyse echter feitelijk onmogelijk. Daarnaast zijn er nog zoveel andere problemen bij de automatische zinsontleding (zie Bakker (1989)), dat het voorlopig nog onmogelijk is om de prosodische zinsstructuur langs deze weg af te leiden.

3 Opzet van de prosodische analyse

De problemen die hierboven aangegeven werden, maken een prosodische analyse nog meteen onmogelijk. De bedoeling is immers om een structuur als in Figuur 1 aan te brengen bij (Nederlandse) zinnen; de manier waarop dat gebeurt, is van minder belang. We nemen daarom onze toevlucht tot een tweede mogelijkheid om de prosodische zinsstructuur te bepalen. Hierbij proberen we de syntactische analyse tot een minimum te beperken. We leiden de prosodische zinsstructuur *rechtstreeks* af uit een zin, zonder de moeizame omweg langs de syntactische analyse. We moeten ons dan wel realiseren dat deze methode slechts een *benadering* oplevert van de prosodische zinsstructuur. In veel gevallen zal de resulterende structuur niet optimaal zijn, omdat we bij de prosodische analyse nu niet kunnen beschikken over alle relevante syntactische informatie.

We weten dat er negen planeten zijn in ons zonnestelsel. Bij de bepaling van de positie van Mars in de sterrenhemel kunnen we de bewegingen van Pluto echter verwaarlozen. Onze taak wordt een stuk eenvoudiger, indien we het aantal factoren in ons model verkleinen (bijv. door de onderlinge aantrekkingskrachten tussen de planeten te negeren). Toch moeten we, in het geval van Mars, soms rekening houden met de aantrekkingskracht van Jupiter. Iets dergelijks geldt ook bij de prosodische zinsanalyse; soms kunnen we niet zonder syntactische informatie omtrent de woorden en woordgroepen van een zin. De alternatieve methode dient dus – in sommige gevallen – een gedeeltelijke syntactische zinsontleding uit te voeren. Bepaalde grammaticale informatie is nu eenmaal essentieel bij de prosodische analyse. Zo moet het hoofdwerkwoord bekend zijn, bijv. teneinde aparte Phi-constituenten te kunnen afbakenen voor onderwerp (subject) en lijdend voorwerp (object).

Deze benadering gaat ervan uit dat de Phi-constituenten grotendeels bepaald kunnen worden op grond van het onderscheid CW/FW, zonder rekening te houden met de syntactische woordsoort. Van elk woord wordt onderzocht of het voorkomt in het lexicon van functiewoorden (omvang ca. 500 woorden); zo niet dan is het een inhoudswoord. Een tweede uitgangspunt is dat in het Nederlands kernen doorgaans uiterst 'rechts' in de constituent staan (zie 3). De eerder vermelde omzettingsregel kan nu *nagebootst* worden met een simpeler regel die uiteindelijk hetzelfde resultaat oplevert. We brengen een Phi-constituentgrens (Phi-Grens) aan tussen elk CW (kern van 'linker' constituent) en daaropvolgend FW (begin van 'rechter' constituent). In taalkundige notatie:

9 → PhiGrens / CW __ FW

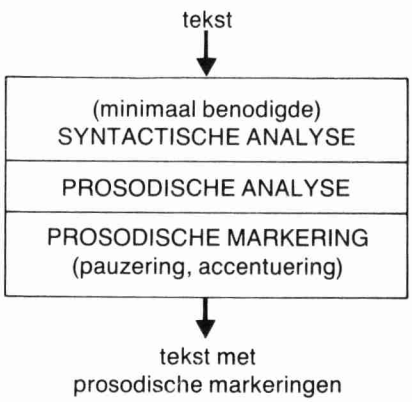
INFORMATIE JAARGANG 31 NR. 7/8
PAG. 559 T/M 636 1989

Deze regel leidt eveneens tot de Phi-constituenten als in Figuur 1, maar zonder de moeizame syntactische analyse. Zo wordt bijv. de kern van een constituent niet expliciet geïdentificeerd. Hoewel deze methode om Phi-constituenten af te bakenen in de praktijk redelijk voldoet, resulteert dit ook wel in foute analyses:

10 (dat de minister verstandig) | (is)
FW FW CW CW | FW

We kiezen dus voor een pragmatische aanpak bij de prosodische analyse en de bepaling van de prosodische spraakkenmerken. De hier geschetste benadering maakt gebruik van absolute (taalkundige) wetten, alsook van statistische wetmatigheden (zoals 9). Ook spellingskenmerken zoals interpunctie, hoofdletters en woordlengte spelen een rol: ze verschaffen immers informatie over hoe een zin uitgesproken dient te worden.

Een en ander impliceert, dat een algoritme voor prosodische zinsontleding drie deeltaken moet verrichten: 1 (minimaal) noodzakelijke syntactische informatie bepalen, 2 feitelijke prosodische analyse uitvoeren, 3 prosodische spraakkenmerken aanbrengen in de uitgevoerde synthetische spraak. Dit is schematisch weergegeven in Figuur 2.



Figuur 2. Belangrijkste modules in de prosodische analyse ten behoeve van een spraaksynthesesysteem

De eerste module van ons algoritme bepaalt woordsoorten (syntactische categorieën) van die woorden die niet te vinden zijn in het functiewoordenlexicon (oftewel: van inhoudswoorden), echter uitsluitend voor zover die van belang is bij de verdere prosodische analyse. Werkwoorden zijn zeer belangrijk voor de prosodische zinsstructuur, en moeten daarom door deze module herkend worden. Dat gebeurt enerzijds met behulp van vormkenmerken (zoals *ge*en, *tte),* en anderzijds met behulp van contextuele restricties (zoals: 'werkwoord mag niet voorkomen na lidwoord'). Verder herkent deze module telwoorden als zodanig en wordt een selectie gemaakt uit de verschillende mogelijke ('dubbele') woordsoorten van sommige functiewoorden (zoals *die, dat).*

Deze eerste module zou deels vervangen kunnen worden door een lexicon ten behoeve van efficiënte grafeem-foneemomzetting (zie Berendsen, Lammens en Van Leeuwen, (1989)), waarin de klankvormen van woorden liggen opgeslagen. Zo'n lexicon kan daarnaast immers ook de syntactische categorie (woordsoort) van elk opgenomen woord bevatten. Deze syntactische categorieën hoeven dan niet uitsluitend via regels afgeleid te worden, maar kunnen ook in dit lexicon opgezocht worden. Het bestand dient dan, naast de functiewoorden, ook de meest voorkomende inhoudswoorden te bevatten. Ook in dit geval zullen sommige regels uit de eerste module toch nodig blijven, bijv. om een selectie te maken uit de 'dubbele woordsoorten' die het lexicon kan specificeren.

De volgende module bakent prosodische constituenten af, door prosodische grenzen te inserteren van verschillende zwaarte (PhiGrens en IntGrens). Het invoeren van grenzen gebeurt op basis van de eerder aangebrachte woordclassificatie als CW of FW (vergelijk regel 9). Daarnaast speelt de syntactische woordsoort een rol bij het afbakenen van werkwoordsgroepen als afzonderlijke Phi-constituent.

De laatste module bevat regels die de prosodische zinsstructuur 'omzetten' in twee soorten prosodische spraakkenmerken: accentuering en pauzering. De regels voor pauzering zijn simpel: maak een pauze (250 ms stilte) bij elke IntGrens in de prosodische zinsstructuur. Het gevolg van deze regel is dus, dat Int-constituenten door pauzes afgebakend worden.

De accentuering verloopt wat minder simpel. In menselijke spraakproduktie wordt accentuering in hoofdzaak bepaald door a de syntactische zinsstructuur en b semantische en pragmatische factoren (Baart (1987); Gussenhoven (1984); zie Kager (1988) voor een samenvatting van de literatuur). Ons model van de menselijke spraakproduktie kan voor accentuering volgens a de syntactische zinsstructuur *nabootsen* door gebruik te maken van de beschikbare prosodische zinsstructuur. In iedere Phi-constituent wordt één woord geaccentueerd (*nucleair* accent), eventueel voorafgegaan door optionele accenten. Functiewoorden zijn in het algemeen niet accentueerbaar, maar partikels van scheidbare werkwoorden weer wel.

11 ik bel mijn moeder nu OP om iets AF te spreken

Toch kunnen we bij de accentuering niet altijd zonder

syntactische informatie. Werkwoorden moeten geïdentificeerd worden, omdat ze normaliter niet geaccentueerd worden. Daarmee vormen ze een uitzondering op bovenstaande regel.
Accentuering volgens b, semantische en pragmatische factoren, is voorlopig vrijwel niet na te bootsen. Waar mogelijk, maken we wel gebruik van pragmatische informatie. Woorden als *dergelijke* en *zulke* leiden meestal iets in dat reeds bekend verondersteld wordt. Die bekende ('gegeven') woorden blijven ongeaccentueerd (12). Constituenten die nieuwe informatie voor de luisteraar bevatten, moeten juist wèl geaccentueerde kernen bevatten. Die nieuwheid van de informatie blijkt vaak uit telwoorden (13), of uit het gebruik van het onbepaald lidwoord (14).

1 Syntactische analyse

<Undef> → <Verb/Part> / (<Prep>) {'ge' } %%* {'t' }
{'be' } {'d' }
{'ver'} {'en'}
{'ont'}

\+ de koning werd *ver*raden
\+ de winkel was *uit*verkocht
– ik vind het *best*

{'die,dat'} → <Pron> / { <Prep> } ___ <CW>
<Comp> <FW> { <Conj> }
\+ in *dat* geval
\+ en *die* man zei
– hij heeft *dat* gezegd

2 Prosodische analyse

→ IntGrens / <VERB> ___ NOT <Verb>
NOT <Flex>
\+ ik vroeg hem te *komen* <IntGrens> *met* al zijn vrienden
\+ zij heeft er in *toegestemd* <IntGrens> *een* plan te schrijven
– hij wil liever *staan* blijven kijken

→ PhiGrens / NOT ZinsGrens <Cap> * ___ NOT <Cap>
NOT *s(e) NOT <grens>
\+ toen Ronald *Reagan* <PhiGrens> *president* was
– dat Hollands*e kaas* lekker is

3 Prosodische markering

<-acc> → <+acc> / *ste
NOT <Verb>
\+ de *MOOISTE* boeken
– hij dans*te* de samba

'Acc-VerbCluster' / ZinsGrens (NOT{<Pron>,<grens>})^n ___ {IntGrn}
{ZinsGrn}
\+ de minister *ONTKENDE* <IntGrens> dat hij op de hoogte was

<+acc> → <-acc> / <KnownQual> ___ (<grens>)
NOT <Pron> NOT{<Verb>,<Num>}
\+ dat hij *zulke boekjes* leest
– dat ik *dit* WIL

Figuur 3: Enkele regels uit het algoritme voor prosodische zinsanalyse, zoals besproken in de tekst, met voorbeelden

12 (ik) (KOCHT) (zulke boeken)
13 (ik) (kocht) (DRIE BOEKEN)
14 (ik) (kocht) (een BOEK)

In figuur 3 laten we enkele regels (in taalkundige notatie) zien. Het effect van deze regels wordt geïllustreerd door de bijgaande voorbeelden.

4 Evaluatie

We hebben ons model van de menselijke spraakproduktie nu uitgebreid met een belangrijke factor: de prosodische zinsstructuur van de uit te spreken zin. Op basis hiervan worden prosodische kenmerken aangebracht in de resulterende synthetische spraak. Voorlopig beperken we die kenmerken overigens tot 1 pauzering en 2 accentuering, zoals besproken in par. 3. We hopen nu dat dit verbeterde model ook inderdaad betere voorspellingen oplevert. De voorspellingen van ons uitgebreide model moeten beter overeenstemmen met de werkelijkheid: de resulterende synthetische spraak moet nu minder verschillen van natuurlijke menselijke spraak dan eerst het geval was.

Deze evaluatie van het spraaksynthesesysteem is op dit moment in volle gang. In luisterexperimenten wordt natuurlijke en synthetische spraak met elkaar vergeleken en worden eventuele perceptieve verschillen tussen die twee vastgesteld. Omdat er nog geen resultaten beschikbaar zijn, zullen we hierop niet verder ingaan.
We kunnen ons uitgebreide model ook nog op een andere manier toetsen. We vergelijken dan niet de perceptie, maar de *produktie* van synthetische en natuurlijke spraak. Het uitgebreide model heeft prosodische spraakkenmerken aangebracht in de uitgevoerde synthetische spraak. We kunnen nu vergelijken in hoeverre een menselijke inspreker dezelfde kenmerken 'aanbrengt' (in dezelfde verzameling zinnen). De menselijke accentuering en pauzering gaan we dus vergelijken met deze spraakkenmerken in de uitvoer van ons spraaksynthesesysteem. In het ideale geval zijn deze identiek: ons model en de menselijke inspreker leveren precies dezelfde uitvoer (althans met betrekking tot deze twee prosodische spraakkenmerken).
De uitkomsten van deze alternatieve evaluatiemethode moeten we overigens wel met een korrel zout nemen. Immers, ook menselijke sprekers maken fouten (zoals we allemaal kunnen vaststellen). Zo kunnen sprekers soms ook foutieve prosodische spraakkenmerken aanbrengen (bijv. accent op reeds bekend woord); dit soort fouten hindert luisteraars aanzienlijk bij hun verwerking van de gesproken boodschap (Nooteboom en Kruyt (1987)). Als we een verschil constateren tussen de uitvoer van ons model en de menselijke spraak, dan impliceert dat dus niet altijd een fout in ons model.
De boven geschetste vergelijking is momenteel alleen uitgevoerd voor wat betreft de produktie van accentuering in natuurlijke en synthetische spraak en nog niet voor de pauzering. Voor deze vergelijking heeft een professionele inspreker een aantal teksten ingesproken op geluidsband (totaal 597 zinnen, 10 734 woorden). In deze teksten (die overigens bedoeld zijn om voorgelezen te worden op de radio) zijn met de hand alle woorden gemarkeerd die door de spreker geaccentueerd werden. Daarnaast zijn door ons algoritme voor prosodische zinsanalyse eveneens accentmarkeringen aangebracht in dezelfde teksten. Het resultaat van de vergelijking tussen deze twee accentmarkeringen is samengevat in Tabel 1.

		inspreker		
		+acc	−acc	totaal
algoritme	+acc	3645 (34.0)	651 (6.1)	4296 (40.0)
	−acc	784 (7.3)	5654 (52.7)	6438 (60.0)
	totaal	4429 (41.3)	6305 (58.7)	10 734 (100)

Tabel 1: Overeenkomsten en verschillen tussen de accentueringen zoals voortgebracht door een menselijke inspreker en bepaald door het algoritme voor prosodische spraakkenmerken; de getallen zijn absolute woordfrequenties, met percentages tussen haakjes. '+acc' resp. '-acc' duidt op wel resp. niet geaccentueerde woorden

Uit deze gegevens kunnen we afleiden, dat ons algoritme bij 86.7% van de woorden (3645+5654) resulteert in dezelfde accentuering als in menselijke spraakproduktie. Deze mate van overeenkomst lijkt vooralsnog zeker acceptabel. Bovendien lijkt de perceptieve kwaliteit van de synthetische spraak (de uitvoer van ons model) aanzienlijk verbeterd te zijn als gevolg van de pauzering en accentuering die we nu aanbrengen op basis van de prosodische zinsstructuur.

Uiteraard wordt er momenteel aan gewerkt om het aantal 'fouten' (discrepanties met de natuurlijk geproduceerde spraak) terug te brengen. Bij deze verbeteringen geldt overigens de wet van de verminderde meeropbrengst. We kunnen een nieuwe regel aan ons algoritme toevoegen, waarmee een gedeelte van de 13.3% 'fouten' correct behandeld wordt. We hopen zo het foutenpercentage terug te dringen. De kans is

echter groot dat deze regel ook (ten onrechte) wordt toegepast bij de woorden die nu reeds correct geaccentueerd worden: toevoeging van een nieuwe regel veroorzaakt weer nieuwe fouten, die weer door een derde regel moeten worden opgelost. enz.
Daarnaast is een gedeelte van de fouten die ons algoritme maakt, principieel niet oplosbaar. Het gaat dan om gevallen waarbij de accentuering volgt uit semantische, pragmatische of contextuele (niet-taalkundige) factoren. De accentuering van de menselijke spreker weerspiegelt dan zijn (impliciete) kennis van de wereld of voegt een nuance toe aan de zin die hij uitspreekt. Een voorbeeld hiervan is de onderstaande zin 15, uitgesproken nadat Van Eekelen moest aftreden als minister van Defensie (naar aanleiding van de paspoortaffaire, september 1988).

15a We hebben weer een GOEDE minister van Defensie.
15b We hebben WEER een goede minister van Defensie.

De twee accentueringen geven hier tegenovergestelde visies weer op de capaciteiten van Van Eekelen. Het politieke standpunt van de spreker blijkt uit zijn accentuering van deze zin (in dit geval was het opvallend dat een partijgenoot van Van Eekelen koos voor de accentuering als in (15a)).
Bij nadere inspectie blijkt 43% van alle fouten die het algoritme maakt, te categoriseren als van semantische, pragmatische of contextuele aard. Zulke fouten zijn alleen op te lossen door uitgebreide kennis van de wereld toe te voegen aan ons algoritme, hetgeen voorlopig nog onmogelijk zal blijven. Het hier beschreven algoritme is vooral taalkundig georiënteerd en zal daarom altijd dit type (niet-taalkundige) accentfouten blijven maken (ca. 6% van de woorden in een tekst).

5 Conclusies

In dit artikel hebben we de spraaksynthese ('sprekende computer') beschouwd als een model van de menselijke spraakproduktie. De uitvoer van ons model kan belangrijk verbeteren, indien we prosodische spraakkenmerken aanbrengen in de synthetische spraak. Vooral accentuering en pauzering vormen voor de luisteraar belangrijke hulpmiddelen bij het achterhalen van de zinsstructuur en dragen daarmee bij tot een beter begrip van de gesproken boodschap. Een spraaksynthesesysteem dient dus een prosodische analyse van de uit te spreken zin uit te voeren. Deze prosodische zinsstructuur van de zin leiden we af met behulp van een relatief simpel algoritme, waarbij de grammaticale ontleding (syntactische analyse) beperkt blijft tot een minimum.

In de toekomst zullen we rapporteren over de definitieve toetsing van ons uitgebreide model: luisterexperimenten met synthetische spraak. Toevoeging van accentuering en pauzering, afgeleid volgens het hier beschreven algoritme, zou moeten leiden tot betere verstaanbaarheid en beter begrip van het gesprokene. Een eerste voorlopige toetsing laat al wel zien dat de automatisch afgeleide accentuering tamelijk goed overeenkomt met die van menselijke spraak. De afwijkingen worden voor een groot deel veroorzaakt door factoren die buiten het blikveld van een taalkundig georiënteerd algoritme liggen. Het aantal uitspraakvarianten van elke boodschap is astronomisch groot, juist met betrekking tot de prosodische spraakkenmerken. Menselijke sprekers maken zeker geen willekeurige selectie uit alle mogelijke uitspraakvarianten. De keuze voor een bepaalde variant is doorgaans zeer adequaat, gegeven factoren als de inhoud van de boodschap, de context waarin deze wordt uitgesproken en de relaties tussen spreker, boodschap en context. De huidige prosodische analyse, zoals geschetst in dit artikel, schiet duidelijk tekort om deze keuze (uit mogelijke prosodische spraakkenmerken) na te bootsen. Deze situatie zal voorlopig niet wezenlijk verbeteren. Immers, geen algoritme voor prosodische zinsanalyse zal ooit in staat zijn om menselijke sprekers te evenaren in hun kennis van de wereld.

Referenties

Baart, J.L.G. (1987), *Focus, syntax, and accent placement: towards a rule system for the derivation of pitch accent patterns in Dutch as spoken by humans and machines.* Dissertatie Rijksuniversiteit Leiden.
Bakker, D. (1989), 'Automatische syntactische analyse'. *Informatie* 31.
Berendsen, E., J. Lammens en H. van Leeuwen (1989), 'Van tekst naar foneemrepresentatie'. *Informatie* 31.
Collier, R. en J. 't Hart (1975), 'The role of intonation in speech perception'. A. Cohen en S.G. Nooteboom (eds.) *Structure and process in speech perception.* Berlijn/Heid./N.Y.; Springer Verlag, blz. 107-21.
Gussenhoven, C. (1984), *On the grammar and semantics of sentence accents.* Dordrecht, Foris.
Kager, R. (1988), 'Plaatsing van zinsaccenten en pauzes in spraak'. M.P.R. van den Broecke (red.) *Ter sprake: spraak als betekenisvol geluid in 36 thematische hoofdstukken.* Dordrecht; Foris, blz. 416-427.
Kaisse, E.M. (1985), *Connected Speech: the interaction of syntax and phonology.* Orlando/London. Academic Press.
Nespor, M. en I. Vogel (1983), 'Prosodic structure above the word'. A. Cutler en D.R. Ladd (red.), *Prosody: models and measurements.* Berlijn/Heidelberg/New York/Tokio, Springer Verlag.
Nespor, M. en I. Vogel (1986), *Prosodic phonology.* Dordrecht, Foris.
Nooteboom, S.G. en J.G. Kruyt (1987), 'Accents, focus distribution, and the perceived distribution of given and new information: an experiment'. *J. Acoustical Soc. America* 82 (5); blz. 1512-24.
Scharpff, P. (1988), 'Persoonlijke communicatie'.
Selkirk, E.O. (1984), *Phonology and syntax. the relation between sound and structure.* Cambridge/London, MIT Press.

De auteurs waren ten tijde van de werkzaamheden waarvan zij in dit artikel verslag doen, werkzaam als toegevoegd onderzoeker bij het SPIN-programma 'Analyse en synthese van spraak' en verbonden aan de Letterenfaculteit van de Rijksuniversiteit Utrecht.