

Dialogen in een Schouwburg-Informatie en Reserveringssysteem

T. Andernach, R. op den Akker, G.F. vd Hoeven, F. de Jong & A. Nijholt
(Universiteit Twente, Faculteit Informatica, Enschede)

S.P. vd Burgt & H. Kloosterman
(KPN Research, Leidschendam)

Abstract

SCHISMA is een onderzoeksproject van de Universiteit Twente en KPN Research. Het project richt zich op de ontwikkeling van een systeem dat een dialoog kan aangaan met een gebruiker die informatie wil over schouwburgvoorstellingen en concerten en daarna eventueel een reservering wil maken. De dialoog gebeurt in natuurlijke taal. Het incrementeel te ontwikkelen systeem dient als een testbed voor de ontwikkeling van de verschillende taalmodulen nodig voor dialoogsysteem. In dit artikel ligt de nadruk, zoals in het gehele project, op het modelleren van de dialogen.

1. Inleiding

SCHISMA (SCHouwburg Informatie Systeem) is een software systeem dat in staat moet zijn de rol te spelen van een informatrice die informatie geeft over voorstellingen of concerten in een bepaalde stad of streek en toestaat voor een voorstelling of concert een reservering te maken. Onderzoek dat moet leiden tot een dergelijk systeem wordt gedaan in de overtuiging dat het niet zozeer meer efficiënte ontleedmethoden of meer volledige methoden voor semantische analyse zijn, maar dat vooral de mogelijkheid tot een vriendelijke, efficiënte en coöperatieve dialoog de sleutel is tot commerciële toepassingen. Dialoogmodellering en dialoogmanagement staan daarom centraal binnen dit onderzoeksproject.

Verschillende onderzoekers (Dahlback 1995; Zoltan-Ford 1991) hebben aangegeven dat mensen hun taalgebruik aanpassen op grond van hun dialoogpartners. Dit geldt niet alleen voor menselijke dialoogpartners, maar ook als gesproken wordt tegen een computersysteem. Een consequentie is dat bij het kijken naar realistische dialogen men niet kan vertrouwen op opnamen of transcripties van mens-tot-mens dialogen. Een bekende manier om dit probleem op te lossen is het gebruik van Wizard-of-Oz (WoZ) experimenten om meer accuraat materiaal te verzamelen. In een WoZ-experiment communiceert de menselijke dialoogpartner (de gebruiker) met een computerprogramma (de WoZ-omgeving). De gebruiker weet niet dat dit 'programma' eigenlijk wordt gestuurd door een onderzoeker die, geheel of gedeeltelijk, de gebruikersuitingen interpreteert en de toepassing in de juiste richting stuurt.

We gebruiken de volgende procedure voor de ontwikkeling van een dialoogsysteem, een combinatie van een procedure voorgesteld door Fraser (1995) en een 'system in the loop' benadering:

- Bestudeer de mens-mens dialogen in het domein waarvoor het systeem ontwikkeld moet worden.
- Definieer WoZ-experimenten.
- Voer de WoZ-experimenten uit.
- Maak transcripties van de verkregen dialogen.

- Specificeer het dialoogsysteem.
- Implementeer een versie van het dialoogsysteem.
- Test het systeem (ontwikkelaars).
- Test het systeem (echte gebruikers)
- Gebruik data van 7 en 8 om het systeem te verbeteren.
- Als teveel modificaties nodig zijn, voer nieuwe WoZ-experimenten uit.
- Ga terug naar stap 6.

Het SCHISMA systeem is een zich uitbreidend 'system-in-the-loop' systeem. Het begon als een puur, door een onderzoeker bemand, WoZ-systeem. In de loop der tijd werden verschillende toevoegingen gemaakt, naarmate door het onderzoek meer subsystemen beschikbaar kwamen. Op dit moment (1996) zijn modulen toegevoegd die betrekking hebben op de analyse van de input (morfolologische analyse, foutcorrectie, identificatie van belangrijke concepten zoals tijd, locatie, naam voorstelling, etc.) en worden modulen toegevoegd die betrekking hebben op dialoogmanagement (wat voor uiting wordt verwacht door systeem of gebruiker).

Vergeleken met andere onderzoekssystemen waarbij een gebruiker de mogelijkheid heeft om via natuurlijke taal vragen te stellen is SCHISMA een uiterst ambitieus systeem. Het heeft meerdere taken (het verschaffen van informatie op een heel breed terrein en het uitvoeren van reserveringen) en het laat een 'spontane' dialoog toe. Daar staat tegenover dat binnen SCHISMA wij ons concentreren op taal en dialoog. Spraakherkenning is een onderwerp dat we in eerste instantie overlaten aan gespecialiseerde onderzoekscentra.

Desalniettemin zijn er tal van projecten die elementen bevatten die een leerzame vergelijking met SCHISMA toelaten. Philips Aken heeft een systeem ontwikkeld voor het telefonisch opvragen van informatie over treinreizen, een via spraak toegankelijke 'Reisplanner'. Het systeem gaat uit van een uiterst simpel (en uiterst effectief) dialoogmodel. De beperktheid van het domein en de doelen van de gebruiker worden daarbij uitgebuit (Aust, 1995). Binnen het NWO Prioriteitsprogramma Taal- en Spraak-Technologie wordt in samenwerking met de OVIS (Openbaar Vervoer Informatie Systeem) organisatie geprobeerd een Nederlandse versie van dit systeem te maken. Er zijn ook tal van systemen in ontwikkeling waarbij geprobeerd wordt met gebruikers een dialoog aan te gaan over het plannen en regelen van een reis per trein of per vliegtuig (Zue et al, 1991, Allen et al, 1994, etc.).

In de volgende secties zal aandacht besteed worden aan de Wizard-of-Oz-experimenten, de architectuur van het SCHISMA systeem, de analyse van de gebruikersuitingen, de analyse en modellering van dialogen en de verwachte ontwikkelingen in het onderzoek en het te realiseren systeem.

2. De Wizard-of-Oz Experimenten

Voor het ontwikkelen van een gebruikersvriendelijk dialoogsysteem is het noodzakelijk om inzicht te verkrijgen in de communicatie tussen mensen en computers; het talig gedrag van mensen tegenover computers verschilt namelijk in verschillende opzichten nogal van dat van mensen onderling. Daarom zal in een zo vroeg mogelijk stadium een beeld verkregen moeten worden van hoe mensen met computers in natuurlijke taal communiceren. Het verzamelen van 'natuurlijke' mens-computer dialogen is in dit stadium van het onderzoek nog niet mogelijk; het uiteindelijke doel is immers juist het ontwikkelen van een systeem waarmee zo'n natuurlijke dialoog kan worden gevoerd. Door het vroege tijdstip van evaluatie moeten bepaalde taken worden gesimuleerd. Een van de mogelijke simulatiemethoden is een zogenaamd Wizard-of-Oz experiment.

In dit soort experimenten zijn de proefpersonen in de veronderstelling dat ze communiceren met de computer terwijl in werkelijkheid verschillende taken worden uitgevoerd door een van de onderzoekers, de wizard. Taken die (al dan niet gedeeltelijk) kunnen worden gesimuleerd zijn onder andere de interpretatie van de uitingen van de proefpersoon, het raadplegen van de database, het interpreteren van de output van de database-query en het kiezen van een vervolguiting.

Het is wenselijk dat de functionaliteit van de Wizard-of-Oz-dialogen dezelfde is als die van de dialogen in het uiteindelijke systeem, omdat met de simulatie beoogd wordt de werking van het uiteindelijke systeem zo dicht mogelijk te benaderen. Dit houdt onder andere in dat de omvang en de complexiteit van de gebruikte database niet te veel mag afwijken van de database van het beoogde systeem en dat de proefpersonen en de taken die zij uit moeten voeren een afspiegeling moeten zijn van de uiteindelijke taken en gebruikers van het systeem.

Verder is het noodzakelijk dat de proefpersonen niet op de hoogte zijn van de aanwezigheid van een wizard, omdat dat het verloop van de dialogen op een ongewenste manier kan beïnvloeden. Een van de punten waar in dit verband dan ook op moet worden gelet is het zo 'computer-like' mogelijk gedrag van het simulatiesysteem; er moet zo consequent mogelijk worden gereageerd op uitingen van de proefpersoon.

In SCHISMA is een simulatie-omgeving ontwikkeld om de complexe taak van de wizard te verlichten; in verschillende vensters heeft de wizard een overzicht van het verloop van de dialoog, de mogelijke vervolguitingen en de inhoud van de database. De database met schouwburginformatie is volledig automatisch toegankelijk en de wizard heeft de beschikking over standaarduitingen (vragen en antwoorden) en over templates: uitingen met open plaatsen die opgevuld kunnen worden met informatie die op dat moment het resultaat is van laatste database-query.

De simulatie-omgeving bestaat uit drie onderdelen: de gebruikersinterface, de wizard-interface en een onderdeel dat zorgt voor de controle van de interactie en het vastleggen van de dialogen. De uitingen worden voorzien van tijdmarkeringen. Wanneer twee cliënt-uitingen niet in de tijd gescheiden zijn door een wizarduiting dan worden deze tot dezelfde beurt gerekend.

De simulatie-omgeving wordt momenteel uitgebreid met een dialoogmodel en een component die cliënt-invoer voorbereidt alvorens deze aan de wizard te presenteren. De laatstgenoemde component zorgt voor morfologische analyse en foutcorrectie en herkent uitingen volgens een lijst van domein-afhankelijke gegevens, zoals namen die in de database voorkomen en datum- en tijdaanduidingen. Bovendien worden woorden opgezocht in een beperkt lexicon en voorzien van de daarin aanwezige lexicale informatie. Op deze wijze functioneert de simulatie-omgeving als een eerste versie van het beoogde geautomatiseerde systeem.

De Sessies

Iedere dialoogsessie bestaat uit drie fasen. In de *eerste fase* wordt proefpersonen gevraagd om deel te nemen aan het experiment. Als dat het geval is wordt hen een instructieformulier gegeven dat vooraf moet worden doorgelezen. Op dit formulier zijn ook vijf mogelijke scenario's vermeld waarvan de proefpersoon er een moet volgen bij het testen van het systeem. Scenario's zijn beschrijvingen van de aard van de informatie of acties die de proefpersoon te weten moet komen of uit moet voeren. Bij het gebruik van scenario's moet ervoor gewaakt worden dat de proefpersonen niet te veel worden gestuurd; het zou kunnen gebeuren dat proefpersonen hun 'query' half oplezend van hun scenario in een uiting omzetten, terwijl dat normaal zelden op die manier gebeurt. Verder worden nog wat algemene gegevens gevraagd zoals leeftijd geslacht en het gebruikte scenario. De proefpersonen waren Informatica-studenten en studenten Toegepaste Onderwijskunde waarvan 69% mannen en 31% vrouwen. De gemiddelde leeftijd lag op 22 jaar.

In de *tweede fase* wordt het systeem geactiveerd door de wizard en verschijnt er een eerste melding op het scherm van de proefpersoon. Aan de daaropvolgende dialoog is verder geen protocol opgelegd; de proefpersoon is vrij in de keuze van zijn uitingen en ook de volgorde van de uitingen van de wizard ligt niet vast. Een gemiddelde dialoog verloopt globaal als volgt: eerst heeft de proefpersoon het initiatief; hij informeert naar de mogelijkheden van een door hem gepland bezoek aan de schouwburg en de gegevens die hij hierbij geeft worden in een query naar de database gestuurd. Als hij vervolgens aangeeft dat hij kaartjes wil reserveren, neemt de wizard het initiatief over en probeert de gegevens die nodig zijn voor de reservering van de proefpersoon te weten te komen. Als de reservering gelukt is volgt er nog een afsluitende subdialoog waarin wordt bedankt en gegroet. De hele dialoog wordt automatisch in logfiles opgeslagen. Ondanks het feit dat in de instructie duidelijk is aangegeven dat proefpersonen van een realistische situatie uit moesten gaan, lijken sommige proefpersonen zich ook te hebben laten leiden door het motief te onderzoeken welke mogelijkheden en beperkingen het systeem had. De samenstelling van de steekproef is hier waarschijnlijk ook een van de oorzaken van. In totaal zijn er 64 dialoogsessies gehouden. Iedere sessie bestond uit gemiddeld 28 beurten (14 beurten per spreker) en duurde gemiddeld 11 minuten. Hieronder volgt een voorbeeld van een sessie (K is de klant, W de wizard):

K: Wat gebeurt er komend weekend (19 maart) in de schouwburg?

W: Op deze datum is er geen uitvoering.

K: En op 19 maart? W: In de periode 18 maart 1994 t/m 18 maart 1994 kunt U naar Deelder Denkt en Indonesische verhalen.

K: En op 20 maart

W: Op deze datum is er geen uitvoering.

K: Hoe laat begint Deelder?

W: De voorstelling begint om 20:00.

K: Wat kost het en zijn er nog plaatsen?

W: Heeft u een reductiekaart?

K: Nee

W: De prijs van de voorstelling "Deelder Denkt" is f26,00. En er zijn nog 82 plaatsen vrij.

K: Ik wil graag 4 plaatsen, bij voorkeur op de eerste rij.

W: U kunt kiezen uit de rangen 1e balkon 1 en zaal 1.

K: Is zaal 1 bij het podium? W: Ja. K: Dan graag in zaal 1 op de eerste rij. W: De kosten voor deze kaartjes zijn f 104,00.

K: Kan ik met Mastercard betalen?

W: Sorry, ik heb alleen informatie over schouwburgvoorstellingen. Informatie over betalingen kunt U het beste telefonisch afhandelen ik zal U het adres en telefoon nummer geven.

W: Naam : Twentse Schouwburg

Adres : Langestraat 49

Postcode : 7511 HB

Woonplaats : Enschede

Telefoon. : 053-858500

Informatie.: 053-858500.

K: Kan ik al een reservering maken?

W: Dat heeft U reeds gedaan!

K: Bedankt

W: Bedankt voor de moeite en tot ziens!

In de *derde fase* vindt er een evaluatie plaats van de dialoogsessie. Deze bestaat uit een aantal meerkeuze-vragen en een aantal daaraan verbonden open vragen. De volgende meerkeuze-vragen worden gesteld:

- Hoe vaak maakt U gebruik van computers? (1 = minder dan 5 uur per week, 5 = meer dan 30 uur per week)

- Had U de indruk dat U een dialoog met een COMPUTER aan het voeren was? (1 = Ja, 2 = Nee)
- Hoe tevreden bent U over de wijze waarop de dialoog met dit systeem is verlopen? (1 = ontevreden, 5 = tevreden)
- Waaraan geeft U de voorkeur voor het verkrijgen van schouwburginformatie, een computer of een mens? (1= Computer, 2 = Mens)
- Als u de voorkeur geeft aan een mens: Zou een aanpassing van het systeem uw voorkeur nog kunnen veranderen? (1 = Ja , 2 = Nee)
- Hoe staat U tegenover het gebruik van computers voor het verkrijgen van dit soort informatie? (1 = Negatief, 5 = Positief)

Uit de evaluatie bleek dat 68% van het aantal proefpersonen meer dan 20 uur per week gebruik maakt van de computer. 73% bleek tevreden over de wijze waarop de dialoog verliep. Men had over het algemeen (76%) de indruk met een computer te communiceren. Vraag 2 bleek nogal dubbelzinnig te zijn: hij was bedoeld om er achter te komen of mensen door hadden dat het systeem werd gesimuleerd. Een ontkenning van deze vraag werd vooral gemotiveerd met argumenten voor het feit dat de communicatie niet vloeiend (als een normale dialoog) verliep. De vraag werd dus meestal geïnterpreteerd als een vraag met de nadruk op 'dialoog' in plaats van op 'computer'. 41% van de proefpersonen gaf de voorkeur aan een mens voor het verkrijgen van schouwburginformatie, maar van die mensen vond 59% dat een aanpassing van het systeem (bijvoorbeeld door het sneller te maken) hun mening hierover kon veranderen. Tenslotte oordeelde 83% positief over het gebruik van computers voor het verkrijgen van dit soort informatie. Opvallend is dat bij de open vragen veelvuldig verwezen werd naar de positief ervaren 'anonimiteit'. Men was niet bang, zoals wel gevoeld werd bij een menselijke informatrice, om domme vragen te stellen of te veel tijd in beslag te nemen.

3. Analyse van de Cliënt-uiting

Uit het karakterrijtje dat de cliënt intikt (of het akoestisch signaal dat deze produceert bij gesproken invoer) tijdens de spreekbeurt moet de door de cliënt bedoelde uiting worden gedestilleerd. Aan een uiting onderkennen we verschillende aspecten.

Allereerst de functie van de uiting in de dialoog. Gaat het om een vraag, en dan om wat voor soort vraag. Of betreft het een antwoord op een vraag. Of gaat het om een uiting die primair de bedoeling heeft de dialoog te sturen. Vaak heeft een dialoog-uiting meerdere functies. Zo heeft ook een vraag de functie de dialoog te sturen. Vervolgens welke informatie wordt door de uiting overgebracht. Dit kan informatie zijn over de cliënt, een bepaalde zaak betreffende het domein, over de dialoog zelf, maar ook over de relatie van de cliënt tot een van deze onderwerpen. Tenslotte, uit welke betekenisvolle elementen bestaat het karakterrijtje? Dat kunnen woorden zijn of combinaties van woorden, of symbolen of symboolrijtjes die staan voor datum-aanduidingen, prijs-aanduidingen of leestekens.

Gebruikers van een dialoogsysteem maken tikfouten en spelfouten, en houden zich niet aan taalregels. Woorden en zelfs complete uitingen zijn op zich, buiten de kontekst van de dialoog, beschouwd, meestal cryptisch of dubbelzinnig. Om uit het karakterrijtje de bedoelde uiting van de cliënt te halen is dan ook kennis nodig van zowel de taal, het domein als van dialoog. In het ideale geval kan dan ook het beste deze kennis gebruikt worden bij het onderkennen van de betekenisvolle elementen in het karakterrijtje.

In de huidige architectuur van SCHISMA-systeem wordt dat niet gedaan. Daarin wordt de morfologische analyse en de foutcorrectie en het verder analyseren van de invoer door de parser uitgevoerd onafhankelijk van de feitelijke toestand van de dialoog. Zo wordt er dus bij de analyse van de invoer geen gebruik gemaakt van het feit dat bijvoorbeeld de laatste uiting van het systeem

een vraag was naar het aantal kaartjes, waardoor de karakterrij 'een' zowel als een lidwoord als een getalaanduiding gelezen wordt. Daardoor zal ook het fout ingetikte rijtje 'en' mogelijk niet als het bedoelde 'een' worden gelezen. Iets wat bij de kennis-gebaseerde aanpak wel het geval kan zijn. Nu is het zo dat de dialoogmanager pas in een later stadium bij de analyse van de invoer op basis van de dialoogkennis, (wat voor reactie wordt op basis van de dialoog-toestand verwacht?) de juiste lezing uit een veld van mogelijke lezingen moet selecteren.

Iedere poging alle mogelijke cliënt-uitingen d.m.v. syntactische regels vast te leggen op zodanige wijze dat aan een analyse van de uiting op basis van die regels een functionele betekenis gekoppeld kan worden, lijkt gedoemd te mislukken. Zijn de regels te streng dan is de kans groot dat invoer die voor ons betekenisvol is niet geanalyseerd kan worden, en zijn ze te algemeen dan levert puur syntactische analyse teveel mogelijke lezingen van de invoer op. Analyse van de invoer zal dan ook gebaseerd moeten zijn op zowel syntactische taalkennis als op semantische kennis die betrekking heeft op het domein.

De MAF-module levert op basis van een lexicon met een beperkt aantal woorden en een lijst van namen van auteurs en voorstellingen en dergelijke, een verzameling van lezingen van de invoer. Een lezing bestaat uit een rijtje items die bestaan uit een woord of woordrijtje, een categorie-aanduidingen, zoals Auteursnaam, Plaatsnaam, Datumaanduiding, Soort-voorstelling, et cetera, en de positie in de invoer.

Bovendien wordt in het item aangegeven hoe goed de matching van het herkende woord met het karakterrijtje is. De lezing wordt vervolgens aangeboden aan de Parser die er op basis van een unificatiegrammatica een semantische feature-structuur, aan toe kent (Akker et al., 1995). Deze structuur kan dan met die van de andere mogelijke lezingen worden aangeboden aan de dialoogmanager, die er een actie aan verbindt.

4. Classificatie van Dialooguitingen

Classificatie van dialooguitingen is nodig omdat de klassen in het uiteindelijke systeem gebruikt kunnen worden voor het voorspellen van acties die het systeem uit moet voeren op een bepaald moment in de dialoog.

Meestal worden subjectieve oordelen gebruikt voor het classificeren van dialooguitingen. In de SCHISMA benadering wordt geprobeerd dat zo veel mogelijk te voorkomen omdat is gebleken dat het voor mensen erg moeilijk is om op basis van intuïtie een klasse-indeling te maken en regels te vinden waarmee die klassen aan uitingen kunnen worden toegekend. Hier is het uitgangspunt dat het classificatie-proces zoveel mogelijk automatisch moet verlopen. Het dialoogcorpus (sectie 2) is daarbij de empirische voedingsbodem.

We gaan er van uit dat er een sterk verband bestaat tussen de vorm van uitingen en hun functie; bij het bepalen van de functie van uitingen speelt de vorm een grote rol. Hierbij moeten we denken aan vormaspecten (cues) als de uitingstype of de syntactische categorie van uitingen; sommige uitingen zijn te vergelijken met zinnen in een tekst en kunnen dus worden getypeerd in termen als declaratief, vragend en gebiedend. Andere uitingen bestaan alleen uit zinsdelen zoals een NP, AP, of PP. Andere cues zijn het type van het subject van uitingen en het type van de al dan niet aanwezige werkwoorden. Ook de aanwezigheid van een vraagteken of specifieke lexicale elementen zoals vraagwoorden en woorden als 'graag' en 'toch' spelen een belangrijke rol. Tabel 1 geeft een overzicht van de cues die gebruikt zijn voor het classificeren van de uitingen in het SCHISMA-corpus.

Cue	Label	Beschrijving van Mogelijke waarden
Spreker	(SP)	systeem of klant

Uitingstype	(UT)	de modus van een zin of de syntactische categorie
Subjecttype	(ST)	syntactische categorie en domein-afhankelijke concepten
Cue Woord	(CW)	woorden die de functie van een uiting veranderen
Type 1ste Werkwoord	(FVT)	syntactische categorie, woorden en domein-afhankelijke concepten
Type 2de Werkwoord	(SVT)	syntactische categorie, woorden en domein-afhankelijke concepten
Vraagteken?	(QM)	aan- of afwezigheid van een vraagteken

Tabel 1. Cues en hun mogelijke waarden

Ter voorbereiding van de automatische classificatie wordt iedere uiting gerepresenteerd door een vector van cues, een cue pattern. Voor het automatisch afleiden van klassen uit deze verzameling van cue patterns is een unsupervised classificatie-algoritme toegepast. Met dit algoritme is het mogelijk om op basis van statistische analyse de meest waarschijnlijke klasseverzameling te vinden m.b.t. de cue patterns. Unsupervised classificatie leverde verschillende mogelijke klasse-indelingen op waarvan er een bestond uit zeven klassen. Een globale beschrijving van de klassen wordt in Tabel 2 gegeven. Klasse 0 en 6 uitingen zonder onderwerp

- uitingen waarin informatie wordt gegeven over domeinobjecten
- vragen over acties of toestanden van spreker of hoorder
- vragen van de cliënt over domeinfeiten
- uitingen waarin informatie wordt gegeven over acties of toestanden van de hoorder
- uitingen waarin informatie wordt gegeven over acties of toestanden van de spreker

Tabel 2. Globale indeling van klassen Unsupervised classificatie leverde ook informatie over de relatieve sterkte van klassen en cues.

Klasse	Relatieve klassesterkte
0	0.596
1	0.705
2	0.167
3	0.067
4	1.000
5	0.331
6	0.199

Tabel 3. Relatieve klassesterkte

Cue	Relatieve invloed
Spreker	0.270
Uitingstype	0.859
Subjecttype	1.000
Type van 1ste werkwoord	0.881
Vraagteken	0.248

Tabel 4. Relatieve invloed van de cues De resultaten laten zien dat sommige klassen sterker zijn dan anderen en dat sommige cues meer invloed hebben dan andere cues. Deze informatie kan worden gebruikt voor het bepalen van de uiteindelijke verzameling klassen en voor het testen met nieuwe sets van cues, waarbij klassen worden samengenomen en cues met weinig invloed worden weggelaten.

Vervolgens is de output van het unsupervised classificatieproces (de cue patterns met hun meest waarschijnlijke klasse) willekeurig opgesplitst in een trainingsset van 75% en een testset van 25%. De trainingsset dient als input voor een supervised classificatie-algoritme waarmee een verzameling regels voor het beschrijven van de relatie tussen de cue patterns en hun klasse kan worden afgeleid. De toepassing van dit algoritme leverde een set van 44 regels op voor het afleiden van de klassen van 206 verschillende cue patterns. Twee van die regels worden hieronder als voorbeeld gepresenteerd.

```
IF SUBJECTTYPE = 2nd pers. Pron.
```

```
AND QUESTIONMARK = yes THEN CLASS = 2 [0 0 96 0 0 0 0]
```

```
IF UTTERANCETYPE = N/NP(s) THEN CLASS = 0 [109 0 0 0 0 0 0]
```

De eerste regel drukt uit dat een uiting met een persoonlijk voornaamwoord in de 2de persoon en een vraagteken in klasse 2 valt. Het is een vraag over een actie of een toestand betreffende de spreker zelf of de hoorder. De tweede regel drukt uit dat alle uitingen die alleen uit een (rijtje van) N of NP bestaan in klasse 0 vallen.

De nauwkeurigheid van de regelset is getest op de testset. Nauwkeurigheid van de regels voor een bepaalde klasse kan worden uitgedrukt door de verhouding tussen het aantal goed voorspelde cue patterns en het totaal aantal cue patterns in die klasse. Tabel 5 toont het aantal cue patterns voor ieder mogelijk paar van feitelijke en voorspelde klassen. De nauwkeurigheid is opgenomen in de laatste kolom. 0

Feitelijk\Voorspeld	0	1	2	3	4	5	6	
0	146	0	0	1	0	0	0	99.3 %
1	3	131	0	0	0	0		97.8 %
2	0	0	75	0	0	0	0	100 %
3	2	0	0	66	0	1	0	95.7 %
4	6	0	0	0	78	2	0	90.7 %
5	1	0	0	0	0	43	0	97.7 %
6	0	0	0	0	0	0	33	100 %

Tabel 5. Feitelijke en voorspelde klassen

Het vinden en afleiden van klassen van uitingen met de hier gebruikte methode is een proces dat in meerdere iteraties moet plaatsvinden; het is noodzakelijk om te trainen en testen met verschillende cues en verschillende klasse-sets. De resultaten tot nu toe zijn veelbelovend; de klassen die door het unsupervised algoritme worden opgeleverd zijn nuttig voor gebruik in een dialoogsysteem. Ook de informatie over het relatieve belang van cues en klassen zoals die door dit algoritme wordt gegenereerd is zeer bruikbaar voor het vinden van de optimale set van klassen en cues. Met het supervised algoritme tenslotte bleek het mogelijk om een voldoende algemene en nauwkeurige regelverzameling af te leiden.

5. De Dialogautomaat

Inleiding

Een deel van het huidige onderzoek concentreert zich op de specificatie van de input-taal voor de dialogmanager. Wat voor dialoggeschiedenis moet de dialogmanager bijhouden en welke

elementen uit de laatste gebruikersuiting zijn nodig om de volgende systeemuiting te produceren? De input-taal moet ruimte bieden ze te beschrijven.

De huidige dialoogmanager binnen SCHISMA is een combinatie van een eindige automaat (EA) en een kontekstlijst, waarbij de toestand van de automaat als het ware de toestand van de dialoog aangeeft, terwijl de kontekstlijst de mogelijkheid biedt om bij een niet direct bij de toestand verwachte gebruikersuiting op zoek te gaan naar een kontekst die wel bij die uiting past.

Eigenschappen van het domein die bij uitstek een rol spelen bij het modelleren zijn:

- Er zijn relatief veel soorten data (vergeleken met andere dialoogsystemen) die relevant kunnen zijn tijdens de dialoog (namen van voorstellingen, voorstellingsdata, namen van auteurs en acteurs, gegevens van de gebruiker, etc.)
-
- Een dialoog bestaat vaak, maar niet altijd uit de volgende fasen:
 1. 'Bladeren'
 2. Specifieke informatie over voorstelling
 3. Reserveren
 4. Bevestiging

Het dialoogsysteem zal deze fasen toelaten (evt. in een andere volgorde) en het zal bij 'specifieke informatie' en vooral bij 'reservering' in interactie met de gebruiker proberen een lijst van gegevens volledig te krijgen. Bij reservering en bevestiging zal het systeem het initiatief moeten hebben.

Schisma is niet het enige dialoogsysteem waarin een EA gebruikt wordt voor het structureren van de dialoog. Bij Verbmobil (Alexandersson, Maier & Reithinger, 1995) bestaat de dialoogmanager uit een EA, een statistische module en een planner. De EA modelleert 'normale' volgorden van (18) dialoogacts. De statistische module (Reithinger & Maier, 1995) wordt gebruikt voor het voorspellen van acts, zodat bij meerdere mogelijkheden de meest waarschijnlijke wordt genomen. De EA is klein en vervult geen belangrijke rol. In een nieuwe versie van het systeem wordt hij niet meer gebruikt. Bennacef, Néel & Maynard (1995) hebben een ATIS (Air Travel Information System) dialoogmodel ontwikkeld op basis van een EA. Ook deze EA met 13 verschillende dialoogacts is klein. Bij de automaat wordt ook een semantisch frame (met slots voor aankomst, vertrek, dag en maand) gebruikt om te beslissen wat de volgende actie van het systeem zal zijn. Ook in het Deense Dialoog Project (DDP, Bækgaard et al., 1995) wordt gewerkt aan een ATIS systeem met een EA ('flowcharts') waarbij de status van informatie-items (volledig bekend, gedeeltelijk bekend, onbekend, nog niet bevestigd, etc.) bepaalt welke actie (bv., 'vraag gebruiker om vollediger informatie') uitgevoerd moet worden.

De EA in SCHISMA

De EA in de dialoogmanager van SCHISMA is te vergelijken met die in DDP (zie boven). Ze is niet gebaseerd op het gebruik van 'speech/dialoog' acts, maar vooral op de inhoud van de uitingen in de dialoog. Echter, in tegenstelling tot DDP geeft deze EA de gebruiker veel vrijheid in het geven van informatie. Het type van de uiting van de gebruiker en de relevante informatie in de uiting worden gecodeerd in de input van de dialoogmanager. Op dit moment zijn de typen nog niet gebaseerd op de resultaten van het onderzoek gerapporteerd in sectie 4. Er wordt gebruik gemaakt van de typen WH-question, yes/no question, request for information and information supply. De relevante inhoud van een uiting komt terecht in informatie-items: gebruikersinformatie-items (informatie over de gebruiker zoals die zichtbaar wordt in de dialoog) en database-informatie-items (entiteiten en attributen van de database). De EA is gebaseerd op toestanden die weergeven welke informatie wel of niet, eventueel gedeeltelijk, bekend is aan het systeem. Transitie van het systeem worden dus gerepresenteerd als veranderingen van variabelen die de bekendheid van

bepaalde informatie in de toestanden van het systeem representeren. Voor details betreffende informatie-items, toestanden en transities, zie Bos (1996).

Het Contextmodel

Een van de belangrijkste nadelen van een EA in een dialoogmanager is de rigiditeit: afwijkingen van voorgedefinieerde paden leiden tot problemen. Voorstellen om hier iets aan te doen hebben bijvoorbeeld betrekking op het uitbreiden van de E.A. met een stapel voor het opslaan van dialoogtoestanden, waarbij eventueel teruggekeerd kan worden op een punt in de dialoog waar een andere subdialoog had moeten starten. Geïnspireerd door die benadering wordt binnen SCHISMA geëxperimenteerd met een contextlijst, een geordende lijst van frames (contexten) waarbij ieder frame een (vroegere) context in de dialoog representeert, compleet met de relevante informatie. Steeds wanneer een uiting ontvangen wordt, wordt de lijst van boven naar beneden doorgezocht om een context te vinden die past bij (consistent is met) die uiting. Het frame wordt dan voorzien van de nieuwe informatie en bovenaan de lijst gezet. Daarna wordt een uiting voor het systeem berekend. Bij het zoeken naar een context wordt gezocht op het niet-inconsistent zijn met de informatie in de recente uiting.

Testen en Experimenten

Het testen van de dialoogmanager is geen eenvoudige taak en tijdrovend. Dit komt voor een belangrijk deel doordat op dit moment niet alle onderzoeksresultaten van die vorm zijn dat ze simpel ingepast kunnen worden binnen het systeem. Desondanks is voor een beperkt aantal volledige dialogen uit het corpus gekeken naar wat toepassing van bovenstaande ideeën oplevert. De testdialogen werden, na met de hand te zijn gecodeerd, aangeboden aan een Prolog versie van de dialoogmanager. Dit leidde tot een groot aantal fouten, deels veroorzaakt door het niet goed kunnen interpreteren en typeren van gebruikersuitingen en het niet aanwezig zijn van de informatie in de database, maar voor een deel ook doordat fouten werden gemaakt door het context-selectie mechanisme. Om een voorbeeld te geven, bekijk de volgende dialoog:

K: Waar gaat het stuk 'De Pelikaan' over?

S: {Beschrijving van 'De Pelikaan'}

K: Wanneer speelt Herman Finkers?

S: Herman Finkers speelt niet in 'De Pelikaan'.

In het eerste deel van de dialoog is geen naam van een artiest genoemd, dus het voorkomen van 'Herman Finkers' is niet inconsistent met de vorige context. Een iets andere definitie van context en/of consistent zijn kan dit soort problemen echter oplossen. Dit geldt ook voor het herkennen van een aantal standaard gebruikte manieren om in een gebruikersuiting te refereren naar de huidige of juist een andere context.

6. Toekomstig Onderzoek

Zoals eerder gesteld, hoewel de integratie van het verschillende onderzoek binnen de onderzoeksomgeving Parlevink een belangrijk doel blijft (en waarbij de integratie vorm krijgt door het incrementeel toevoegen van nieuwe onderdelen die leiden tot een meer volledig geautomatiseerd systeem), is het niet zo dat volledig van te voren vastgelegd is welke benaderingen en methoden in het uiteindelijke systeem zullen worden geïntegreerd. SCHISMA is tevens een onderzoeksvehikel om verschillende benaderingen en methoden te kunnen vergelijken en er ervaring mee op te doen. Zo wordt op dit moment onderzoek verricht naar de in sectie 3 genoemde kennisgebaseerde analyse van de gebruikersuiting waarbij ook de toestand van de dialoog betrokken wordt in het komen tot een interpretatie. Wat betreft het classificeren van uitingen wordt o.a. gekeken in hoeverre er gebruik kan worden gemaakt van een (Kohonen) Self-Organizing Feature Map bij het vinden van klassen in uitingen van het SCHISMA-corpus. Wat betreft het dialoogmodel van sectie 5 wordt gedacht aan meer accuraatheid bij het hanteren van het

kontekstmechanisme door gebruik te maken van een meer gedetailleerde typering van gebruikersuitingen en het toevoegen van een typering van (anaforische) referenties. Daarnaast wordt gekeken naar het vervangen van de kontekstlijst door een kontekstboom. Andere benaderingen die systematisch aandacht zullen krijgen zijn het toevoegen van statistische informatie, topic-focus onderzoek en modellering gebruikmakend van 'update' semantiek.

7. Referenties

Akker, R. op den, H. ter Doest, M. Moll & A. Nijholt (1995). Parsing in dialogue systems using typed feature structures. In: *Proceedings 3rd International Workshop on Parsing Technologies (IWPT'95)*, Prague, Karlovy Vary, 1995, 10-11.

Alexandersson, J., E. Maier & N. Reithinger (1995). A robust and efficient three-layered dialogue component for a speech-to-speech translation system. *Proceedings of the 33rd ACL Conference*, Boston, MA.

Allen, J.F. en L.K. Schubert et al. (1994). The TRAINS Project: A case study in defining a conversational planning agent. Technical Report TR 532, URCSD.

Aust, H. en M. Oerder (1995). Dialogue control in automatic inquiry systems. *Proceedings TWLT9: Corpus-based approaches to dialogue modelling*. Enschede, 45-49.

Bækgaard, A., et al. (1995). The Danish spoken language dialogue project. *ESCA Workshop on Spoken Dialogue Systems*, Vigsø, Denmark.

Bennacef, S.K., F. Néel & H.B. Maynard (1995). The Danish spoken language dialogue project. *ESCA Workshop on Spoken Dialogue Systems*, Vigsø, Denmark.

Bos, D.H.R. (1996). Modelling dialogues with finite state automata. Msc. thesis, University of Twente, Enschede, the Netherlands.

Burgt, S. vd, H. Kloosterman, R. Bos & A. Nijholt (1996). *Building dialogue systems that sell. NLP and Industrial Applications*, New Brunswick, June 1996, to appear.

Dahlback, N. (1995). Kinds of agents and types of dialogues. *Proceedings TWLT9: Corpus-based approaches to dialogue modelling*. Enschede, 1-11.

Doest, H. ter, M. Moll, R. Bos, S.P. van de Burgt & A. Nijholt (1996). Language Engineering in Dialogue Systems. *Computers in Engineering Symposium. Session on Natural Language in Human-Computer Interfaces*. Volume I, Book VI, Houston, Texas, February 1996, 68-79.

Fraser, N.M. (1995). Messy data, what can we learn from it? *Proceedings TWLT9: Corpus-based approaches to dialogue modelling*. Enschede, 95-105.

Reithinger, N. & E. Maier (1995). Utilizing statistical dialogue act processing in Verbmobil. Verbmobil Report 80, DFKI, Saarbrücken, Germany.

Zoltan-Ford, E. (1991). How to get people to say and type what computers can understand. *Int. J. Man-Machine Studies* 34, 527-547.

Zue, V., J. Glass, et al. (1991). Integration of speech recognition and natural language processing in the MIT VOYAGER system. *International Conference on Acoustics, Speech and Signal Processing 1991*, IEEE, New York, 713-716.