



Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

Document

D-98-03

**Hybride konnektionistische, statistische
und regelbasierte Ansätze zur Verarbeitung
natürlicher Sprache**

**Workshop auf der 21. Deutschen Jahrestagung für
Künstliche Intelligenz, Freiburg, 9.–10. September 1997**

Stephan Busemann, Karin Harbusch, Stefan Wermter (Hrsg.)

Januar 1998

Deutsches Forschungszentrum für Künstliche Intelligenz

Postfach 20 80
67608 Kaiserslautern, FRG
Tel.: + 49 (631) 205-3211
Fax: + 49 (631) 205-3210
E-Mail: info@dfki.uni-kl.de

Stuhlsatzenhausweg 3
66123 Saarbrücken, FRG
Tel.: + 49 (681) 302-5252
Fax: + 49 (681) 302-5341
E-Mail: info@dfki.de

WWW: <http://www.dfki.de>

Deutsches Forschungszentrum für Künstliche Intelligenz

DFKI GmbH

German Research Center for Artificial Intelligence

Founded in 1988, DFKI today is one of the largest nonprofit contract research institutes in the field of innovative software technology based on Artificial Intelligence (AI) methods. DFKI is focusing on the complete cycle of innovation — from world-class basic research and technology development through leading-edge demonstrators and prototypes to product functions and commercialization.

Based in Kaiserslautern and Saarbrücken, the German Research Center for Artificial Intelligence ranks among the important "Centers of Excellence" worldwide.

An important element of DFKI's mission is to move innovations as quickly as possible from the lab into the marketplace. Only by maintaining research projects at the forefront of science can DFKI have the strength to meet its technology transfer goals.

DFKI has about 115 full-time employees, including 95 research scientists with advanced degrees. There are also around 120 part-time research assistants.

Revenues for DFKI were about 24 million DM in 1997, half from government contract work and half from commercial clients. The annual increase in contracts from commercial clients was greater than 37% during the last three years.

At DFKI, all work is organized in the form of clearly focused research or development projects with planned deliverables, various milestones, and a duration from several months up to three years.

DFKI benefits from interaction with the faculty of the Universities of Saarbrücken and Kaiserslautern and in turn provides opportunities for research and Ph.D. thesis supervision to students from these universities, which have an outstanding reputation in Computer Science.

The key directors of DFKI are Prof. Wolfgang Wahlster (CEO) and Dr. Walter Olthoff (CFO).

DFKI's six research departments are directed by internationally recognized research scientists:

- Information Management and Document Analysis (Director: Prof. A. Dengel)
- Intelligent Visualization and Simulation Systems (Director: Prof. H. Hagen)
- Deduction and Multiagent Systems (Director: Prof. J. Siekmann)
- Programming Systems (Director: Prof. G. Smolka)
- Language Technology (Director: Prof. H. Uszkoreit)
- Intelligent User Interfaces (Director: Prof. W. Wahlster)

In this series, DFKI publishes research reports, technical memos, documents (eg. workshop proceedings), and final project reports. The aim is to make new results, ideas, and software available as quickly as possible.

Prof. Wolfgang Wahlster
Director

Hybride konnektionistische, statistische und regelbasierte Ansätze zur Verarbeitung natürlicher Sprache

Stephan Busemann, Karin Harbusch, Stefan Wermter (Hrsg.)

DFKI-D-98-03

© Deutsches Forschungszentrum für Künstliche Intelligenz 1998

This work may not be copied or reproduced in whole or part for any commercial purpose. Permission to copy in whole or part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of the Deutsche Forschungszentrum für Künstliche Intelligenz, Kaiserslautern, Federal Republic of Germany; an acknowledgement of the authors and individual contributors to the work; all applicable portions of this copyright notice. Copying, reproducing, or republishing for any other purpose shall require a licence with payment of fee to Deutsches Forschungszentrum für Künstliche Intelligenz.

ISSN 0946-0098

**Hybride konnektionistische, statistische
und regelbasierte Ansätze
zur Verarbeitung natürlicher Sprache**

**Workshop auf der 21. Deutschen Jahrestagung
für Künstliche Intelligenz, KI-97
Freiburg, 9.–10. September 1997**

herausgegeben von

Stephan Busemann, DFKI GmbH
Karin Harbusch, Universität Koblenz
Stefan Wermter, Universität Hamburg

DFKI GmbH
Saarbrücken, Januar 1998

Inhalt

Vorwort	iii
Workshop-Programm	v
Beiträge	
Oliver Jokisch, Markus Peschek, Technische Universität Dresden: <i>Perzeptive Akzeptanz neuronal modellierter Intonationskonturen</i>	1
Marc Kupietz, Universität Bielefeld: <i>Lokaler vs. parallel distribuiertes Konnektionismus in der Modellierung der Sprachproduktion</i>	7
Andreas Mertens, Fernuniversität Hagen: <i>Heuristische Ergebnisbewertung beim Wortagentenparsing</i>	17
Hagen Langer, Universität Osnabück, Stephan Mehl, Universität Duisburg, Martin Volk, Universität Zürich: <i>Hybride NLP-Systeme und das Problem der PP-Anbindung</i>	27
Joseph Chen, Stefan Wermter, Universität Hamburg: <i>Initial results on hybrid connectionist bilingual alignment</i>	35
Manfred Stede, Technische Universität Berlin: <i>Identifying dialogue acts: linguistics or statistics?</i>	45
Jens-Uwe Moeller, Universität Hamburg: <i>Dia-MOLE: Learning Spoken Dialogue Models By Integrating Multiple Knowledge Sources</i>	53
Reinhard Blutner, Humboldt-Universität Berlin: <i>Nichtmonotones Schliessen und neuronale Netze</i>	65

Vorwort

Obwohl in den beiden Bereichen der regelbasierten Sprachverarbeitung und der konnektionistischen bzw. statistischen Sprachverarbeitung am selben Forschungsgegenstand gearbeitet wird, sind hybride, einander ergänzende Ansätze erst in jüngster Zeit intensiver untersucht worden. Bisher wurde auf internationaler Ebene versucht, die Vertreter der einzelnen Forschungsansätze miteinander ins Gespräch zu bringen (z.B. auf einem IJCAI-Workshop 1995¹). Einen solchen Brückenschlag, der für beide Verarbeitungsparadigmen vorteilhaft ist, strebte der hier dokumentierte Workshop im nationalen Rahmen an. Die Fachgruppe 1.3.1 "Natürlichsprachliche Systeme" (<http://www.dfki.de/services/gi-nlp/>) der Gesellschaft für Informatik bildete den organisatorischen Rahmen.

Wie stehen nun regelbasierte und konnektionistische oder statistikbasierte Verfahren zueinander? Als generellen Nachteil von regelbasierten Systemen kann man ihre 'Exaktheit' ansehen: fällt das Problem aus dem vordefinierten Rahmen (aus der Domäne), so wird das Ergebnis arbiträr. Deshalb werden heute z.B. Fuzzy-Logiken zur Aufweichung von absoluten Aussagen eingesetzt, und falls die vorhandene Information nicht zur Lösung des Problems ausreicht, lässt sich mittels Unterspezifikation der Eingabe ein (allerdings qualitativ schlechteres) Ergebnis berechnen.

Die konnektionistischen und statistikbasierten Verfahren haben bei diesen Nachteilen der regelbasierten Ansätze ihre Stärken, d.h. sie unterstützen Lernfähigkeit und Robustheit. Andererseits liegen ihre Schwächen bei den Vorteilen regelbasierter Verfahren: es sind keine für den Menschen direkt interpretierbare Regeln vorhanden. Dadurch bleiben die 'Schlüsse' eines solchen Systemes für den Benutzer weitgehend undurchschaubar, sofern nicht zusätzliche Extraktionsmethoden symbolische Interpretationen liefern können.

Auf verschiedenen Repräsentationsebenen können unterschiedliche Ansätze kombiniert oder integriert werden, z.B.:

- Neuronale Netze und regelbasierte Repräsentationen
- Statistische Verfahren und regelbasierte Repräsentationen
- Lokalistische und verteilte konnektionistische Repräsentationen
- Regelbasierte Fuzzy-Repräsentationen und neuronale Repräsentationen

Eine Kombination verschiedener Strategien ist bei vielen linguistischen Verarbeitungsschritten denkbar und wünschenswert, um die Vorteile der verschiedenen Paradigmen an geeigneter Stelle ausnutzen zu können. Hybride Ansätze können für verschiedene Aufgaben sinnvoll sein, insbesondere für Aufgaben, die unterschiedliche Repräsentationen erfordern, z.B.

- Analyse gesprochener Sprache (numerische Signale, symbolische Strukturen)
- Textanalyse (syntaktische, semantische, pragmatische Präferenzen)
- Inferenzbildung (scharfes und unscharfes Schliessen)
- Wissensrepräsentation (regelbasiertes und beispielbasiertes Wissen)

¹S. Wermter, E. Riloff und G. Scheler (Hrsg.): *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*. Springer, Berlin, 1996.

Bei all diesen Beschreibungen stellt sich stets die Frage, wie ein umfassendes hybrides System zur Verarbeitung natürlicher Sprache aufgebaut sein soll. Möglichkeiten hierzu sind im Rahmen dieses Workshops erörtert worden, angeregt durch Fragestellungen wie

- welche existierenden Methoden lassen sich effektiv hybrid kombinieren?
- welche Form der losen, engen Kopplung oder starken Integration ist für eine gegebene Aufgabe sinnvoll?
- welche neuen Methoden sind erforderlich?
- welcher Grad an Modularität oder Interaktivität ist effektiv?
- wie lassen sich hybride mit homogenen Methoden vergleichen?

Die schriftlichen Beiträge in diesem Dokument sind begutachtete unpublizierte Forschungsberichte oder begutachtete Positionspapiere. Ihre Thematik konzentriert sich auf hybride Ansätze in verschiedenen Bereichen der natürlichsprachlichen Verarbeitung. Die Begutachtung erfolgte durch die Organisatoren des Workshops. Von den neun akzeptierten Beiträgen sind acht hier mit dem freundlichen Einverständnis der Autorinnen und Autoren abgedruckt; der Beitrag von Fertig und Scheler, der auf dem Workshop nicht vorgestellt wurde, findet sich auf der Webseite des Workshops unter der URL <http://www.dfki.de/busemann/ki97/ki97-ws03.html>.

Der Workshop wurde von etwa 25 Personen besucht. Es wurde festgestellt, daß sich bisher keine einheitliche Definition für "hybrides System" durchgesetzt hat, denn zu unterschiedlich sind die Kombinationsmöglichkeiten und die damit verfolgten Ziele. Für die deutsche Sprache ergeben sich besondere Schwierigkeiten durch das Fehlen großer annotierter Korpora, die die Performanz statistikbasierter und anderer datengestützter Verfahren bestimmen. Für einige der im folgenden vorgestellten Arbeiten zu hybriden Systemen ist die Nutzung eines annotierten Korpus für das Deutsche erforderlich. Aktuelle Informationen zur Entwicklung deutscher Korpora finden sich unter der URL <http://www.coli.uni-sb.de/cl/projects/negra.html>. Zur Vorgehensweise bei der Annotation von Korpora findet im August 1998 im Rahmen der "European Summer School in Logic, Language and Information" (ESSLLI) in Saarbrücken (<http://www.coli.uni-sb.de/esslli/>) ein Workshop statt.

Stephan Busemann
Karin Harbusch
Stefan Wermter

Workshop-Programm

Dienstag, 9. September 1997

14:00–14:30 **Stefan Wermter**, Universität Hamburg:
*Aktuelle Entwicklungen hybrider Verarbeitung natürlicher Sprache:
eine Einführung*

Block 1: Modellierung gesprochener Sprache

14:30–15:00 **Oliver Jokisch**, **Markus Peschek**, Technische Universität Dresden:
Perzeptive Akzeptanz neuronal modellierter Intonationskonturen

15:00–15:30 **Marc Kupietz**, Universität Bielefeld:
*Lokaler vs. parallel distribuiertes Konnektionismus in der
Modellierung der Sprachproduktion*

15:30–16:00 KAFFEPAUSE

Block 2: Parsing

16:00–16:30 **Andreas Mertens**, Fernuniversität Hagen:
Heuristische Ergebnisbewertung beim Wortagentenparsing

16:30–17:00 **Hagen Langer**, Universität Osnabrück, **Stephan Mehl**,
Universität Duisburg, **Martin Volk**, Universität Zürich:
Hybride NLP-Systeme und das Problem der PP-Anbindung

17:00–17:30 Diskussion

Mittwoch, 10. September 1997

Block 3: Dialog und Alignment

14:00–14:30 **Joseph Chen**, **Stefan Wermter**, Universität Hamburg:
Initial results on hybrid connectionist bilingual alignment

14:30–15:00 **Manfred Stede**, Technische Universität Berlin:
Identifying dialogue acts: linguistics or statistics?

15:00–15:30 **Jens-Uwe Moeller**, Universität Hamburg:
*Dia-MOLE: Learning spoken dialogue models by integrating
multiple knowledge sources*

15:30–16:00 KAFFEPAUSE

Block 4: Semantikepräsentation und Schließen

16:00–16:30 **Niels Fertig**, Universität München, **Gabriele Scheler**,
Technische Universität München:
Constructing semantic representations using the MDL principle

16:30–17:00 **Reinhard Blutner**, Humboldt-Universität Berlin:
Nichtmonotones Schließen und neuronale Netze

17:00–17:30 Schlußfolgerungen – Eine offene Diskussion der Teilnehmer

Perzeptive Akzeptanz neuronal modellierter Intonationskonturen

Oliver Jokisch, Markus Pescheck

Technische Universität Dresden, Institut für Technische Akustik

Kurzfassung

Sprachsynthesysteme verwenden zumeist regelbasierte und überwiegend symbolische Verfahren zur Prosodiegenerierung, welche zur melodischen Monotonie neigen und linguistisch-phonetisches Expertenwissen erfordern. Seit einigen Jahren werden konnektionistische Ansätze z.B. zur Erzeugung der f₀-Kontur publiziert. Die Performanz der verschiedenen Ansätze wird meist entweder mittels Expertenurteil (z.B. über die semantische Korrektheit der resultierenden Äußerungen) oder mittels einer Distanzmessung zu realen Sprachdaten bewertet. Diese unterschiedlichen Bewertungsstrategien erschweren den Vergleich symbolischer, konnektionistischer oder hybrider Verfahren und entsprechen nicht unbedingt den Hörbedürfnissen von Standardnutzern der Sprachausgabesysteme.

Der vorliegende Beitrag diskutiert die Urteile "unvorbelasteter" Hörer. Exemplarisch wird ein neuronaler Ansatz zur f₀-Generierung präsentiert und im Kontrast zu einem regelbasierten Ansatz perzeptiv getestet.

1 Motivation

Am Institut für Technische Akustik der TU Dresden wurde in den zurückliegenden Jahren ein modulares Text-to-Speech (TTS) System [Hir96] für die deutsche Sprache entwickelt. Das System besteht aus der Vorverarbeitung (Graphem-Phonem-Umsetzung, Akzente, Phrasen), dem Prosodiemodul mit Aufbereitung der Syntheseeinheiten und der Akustischen Synthese (der Konkatenation von Diphonen bzw. hybriden Einheiten). Das Dresdner TTS-System bzw. Teilkomponenten kommen in verschiedenen Projekten zum Einsatz - so z.B. in einer PCMCIA-Card für mobile Sprachausgabe [Kor96] oder im Forschungsverbund Verbmobil [VeM95].

Die Kritik der Hörer bezieht sich neben der segmentalen Grundqualität v.a. auf die monotone Intonation, die im Modus "Textvorlesen" auftritt. Dabei wurde ein regelbasiertes, lineares Modell zur f₀-Konturplanung verwendet [Schn94], welches bei ähnlicher Akzent- und Phraseninformation auch zu ähnlichen Ergebnissen neigt.

Inspiziert durch die Arbeiten von Traber [Tra92] versuchen die Autoren, ein alternatives Prosodiemodul auf neuronaler Basis zu entwerfen, welches in das Konzept des Dresdner TTS-Systems paßt. Erste eigene Erfahrungen mit einer neuronalen Komponente zur f₀-Konturplanung [Jok95] und weitere Veröffentlichungen seitens Traber - z.B. [Tra96] - belegen:

- Es ist eine umfangreiche symbolische Modellierung außerhalb des Neuronalen Netzes (NN) notwendig.

- Obwohl die NN-Ergebnisse perzeptiv durchaus ermutigend sind, weichen sie teilweise erheblich vom natürlichen f₀-Kontur-Vorbild ab.

Die durchaus verständliche Skepsis gegenüber neuronalen Lösungsansätzen bewog die Autoren, vor einem Design des Intonationsmoduls zunächst die perzeptive Akzeptanz neuronal erzeugter f₀-Konturen an "unvorbelasteten" Hörern zu testen: Die hohe "Meßlatte" der realen Kontur und damit die Frage nach dem geeigneten Distanzmaß zur Vorgabekontur (Trainingsmaterial) entfällt damit zunächst. Außerdem lassen sich Anregungen gewinnen, die ein einseitiges Verfolgen konnektionistischer Paradigmen verhindern.

In dem Abschnitt "2 Netze kontra Regeln?" werden kurz einige Grundideen zur Prosodiegenerierung kontrastiert. Die Abschnitte "3 Regelbasierte Intonation im TTS-System" und "4 Neuronale Modellierung" beschreiben die im Hörtest verglichenen Ansätze zur f₀-Konturplanung im Überblick. Der Hörtest sowie Ergebnisse werden ausführlich im Abschnitt "5 Hörtest & Ergebnisse" diskutiert und die entsprechenden Schlußfolgerungen im 6. Abschnitt dargestellt.

2 Netze kontra Regeln?

Obwohl die Sprachsynthese zunehmend Einzug in die Praxis hält, wird die synthetische Sprache nach wie vor meist als "unnatürlich", unangenehm, etc. beurteilt. Ein nicht unbedeutender Anteil dieser Vorbehalte ist auf die ungenügende prosodische Modellierung zurückzuführen. Eine Prosodiegenerierung umfaßt die Steuerung der Grundfrequenz f₀, von Sprechrhythmus bzw. Dauer sowie der Lautstärke. Die resultierende Strukturierung und Akzentuierung des Sprechflusses ist einerseits informationstragend, beinhaltet aber auch Aspekte von Natürlichkeit und "Wohlklang" der Sprache. Die regelbasierte Prosodiemodellierung ist folglich komplex:

- umfangreiches Regelwerk
- Kontextabhängigkeiten schwer erfaßbar
- Prosodie variiert stark mit Sprecher, Intention, etc.

Aufgrund dieser Randbedingungen tendieren regelbasierte Prosodiesteuerungen zu melodischer und rhythmischer Monotonie. Die Adaption an verschiedene TTS-Systeme ist in der Regel aufwendig.

Der Gedanke liegt nah, daß eine stärkere Reflexion humaner Informationsverarbeitung, einschließlich Lernalgorithmen und Fuzzy-Ansätzen zu einer höheren Systemperformanz führt. Kernstück könnte ein neuronales Netz sein; mit den bekannten Möglichkeiten:

- automatische Abstraktion von Regularitäten oder Modellparametern
- Generalisierung, Mittelwertbildung
- sinnvolle Ausgaben auch für unbekannte Eingaben
- ...

Leider wirft dieser theoretisch interessante Ansatz neue Probleme auf:

- Generierung und Auswahl der Datenbasis, angemessene Netztopologie
- Welche Modellierung außerhalb des NN ist mindestens erforderlich, welche sinnvoll?

Neuronal motivierte Prosodieansätze sind ergo ähnlich komplex und dementsprechend kritisch zu betrachten.

3 Regelbasierte Intonation (RG) im Dresdner TTS-System

Der derzeitige Implementationsstand des Dresdner TTS-Systems wird in [Hir96] erläutert: Dabei erfolgt die Planung der Grundfrequenzwerte für eine prosodische Manipulation des Synthesesignals anhand eines einfachen Intonationsmodells [Schn94], das von einem stückweise linearisierten f0-Verlauf über die Länge der Phrase ausgeht.

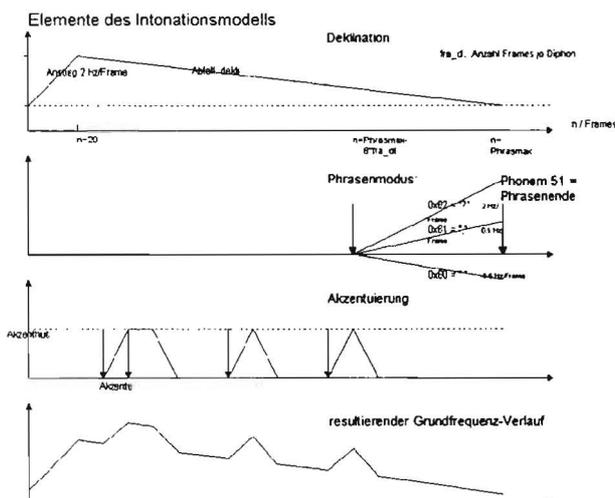


Abb. 1: Intonationsmodell im Überblick aus [Hir96]

Nach dem Vorbild des Fujisaki-Modells [Fuj88] werden Phrasenkomponenten und Wortkomponenten, welche sich aus symbolischer Textinformation ergeben, überlagert. Der Algorithmus für die Phrasenintonation geht

von einem konstanten Grundfrequenzabfall Δf je Phrase aus. Die Phonemanzahl wird auf eine Framezahl an Grundperioden abgebildet. Anfänglich erfolgt ein f0-Anstieg über einer konstanten Framezahl der Phrase. Es werden verschiedene Satzmodi unterschieden, wobei die Zahlenwerte von Baseline, Deklination, Anstieg und Abfall frei konfigurierbar sind, z.B. für eine männliche Stimme:

1. Start: 80 Hz, Frames: 20, Anstieg Δf : 1.7 Hz/Frame
2. Aussage: f0-Abfall von 0.6 Hz/Frame
3. Frage: f0-Anstieg von 2 Hz/Frame
4. Kontinuirung: f0-Anstieg von 0.1 Hz/Frame
- ...

Die Wortintonation erhöht f0 bei Auftreten des entsprechenden Langvokal-Akzenten um einen Wert Akzenthub (z.B. 12 Hz), sofern das vorhergehende Diphon nicht akzentuiert war. Bei "Akzent-Rücksetzen" folgt ein f0-Abfall auf Deklinationsniveau, ansonsten eine Fortsetzung auf höherem Parallelniveau. Eine Akzentuierung von Kurzvokalen bewirkt keine f0-Variation sondern lediglich eine Veränderung der Segmentdauer.

Die Grundelemente des regelbasierten Intonationsmodells und deren Überlagerung zum endgültigen Verlauf veranschaulicht Abbildung 1. Nach erfolgter (regelbasierter) Segmentdauer- und Lautstärkegenerierung werden die Ergebnisse des Prosodiemoduls in ein Steuerfile (SM-File) für die Akustische Synthese eingetragen (Bausteinauswahl, Stützstellen mit relativen f0-Werten, absolute Lautdauern in ms und Lautstärkekommandos in dB). Diese Schnittstelle zur Akustischen Synthese wird auch durch alternative Prosodiemodule verwendet, woraus sich die Möglichkeit des direkten Vergleichs ergibt. Fehlende Stützstellen in der Konturplanung werden im Synthetisator interpoliert.

Für die Akustische Synthese sind entweder ein Diphon-Zeitbereichsverfahren (männliche und weibliche Stimme) oder ein Parallel-Reihen-Formantsynthetisator verwendbar.

4 Neuronale Modellierung (NN)

Die Autoren untersuchten zunächst ein Multilayer Perceptron (MLP) [Jok95]. Als Eingangsmuster diente die kodierte Phonemrepräsentation mit Phrasen- und Akzentsymbolen jeweils einer kompletten Äußerung. Der Ausgangsvektor bestand aus den dazugehörigen, phonemmittigenbezogenen f0-Werten. Innerhalb stimmloser Abschnitte wurde linear interpoliert. Für das Training wurden phonetisch-phonologisch etikettierte (gelabelte) Sätze zur Zugauskunft aus der PhonDat-Datenbank [Pho93] verwendet. Ungeachtet der einfachen MLP-Topologie, des zu geringen Datenumfangs von 43 Sätzen und einer mehr oder minder willkürlichen Auswahl der kodierten phonetischen Attribute bildete dieses Modell

z.B. die Frage- oder Aussageintonation recht gut ab. Nachteilig war die Tendenz zu Über- und Fehlbetonungen und die, in Hinsicht auf die Satzlänge beschränkte, Verarbeitungsbreite der MLP-Eingangsschicht. Die f0-Generierung (Lernphase) auf Basis dieses Ansatzes ist in Abbildung 2 dargestellt.

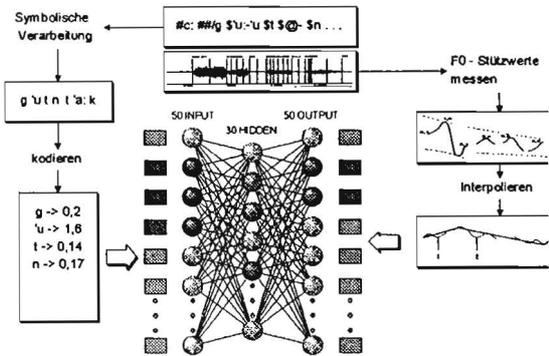


Abb. 2: MLP-Ansatz nach [Jok95]

Ausgehend von einem Ansatz von C. Traber [Tra92], die f0-Konturgenerierung mit einem teilrekurrenten NN zu lösen, wurde ein ähnliches Netz entworfen. Da die von Traber vorgeschlagene symbolische Vorverarbeitung mit entsprechender phonologischer Kodierung innerhalb des Dresdner TTS-Systems nicht möglich ist, wurden Kodierungsschema sowie Netztopologie entsprechend verändert. Abbildung 3 zeigt diesen Ansatz zur f0-Konturgenerierung (Lernphase).

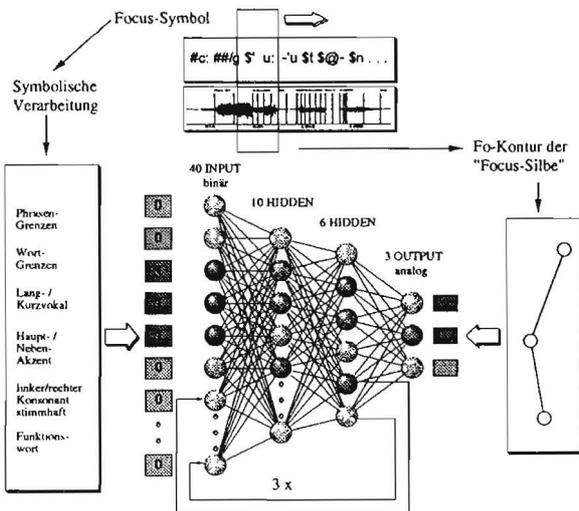


Abb. 3: teilrekurrenter NN-Ansatz nach [Pes97]

Grundidee: Aus der Phonemsequenz der Äußerung werden Pseudosilben (mit jeweils einem Vokal) gebildet, Attribute wie z.B. Funktionswort ja/nein, Haupt-/Nebenakzent, u.ä. abgeleitet und bitweise kodiert (8bit/Silbe). Die silbenorientierte Sequenz von Attributvektoren wird schrittweise an den, aus 5 Silben bestehenden, Netzeingang (2x Vorkontext, 1x Fokussilbe, 2x Nachkon-

text) angelegt. Am Netzausgang werden synchron jeweils 3 f0-Werte je Fokussilbe präsentiert und mit Fehlerminimierung trainiert. Die implementierten Teilrekurrenzen und Kontextzellen sollen die zeitlich-physiologischen Aspekte der f0-Ausgabekontur nachbilden. Ein typisches Kodierungsbeispiel ist in Abbildung 4 dargestellt.

```

#apl ph wb pa sa lv l r key Syllable
-----
file: 5130
10386 1 1 0 0 0 0 0 1 QIC 10752 0.223 11520 0.181 12416 0.208
12870 0 1 0 0 0 0 1 0 1 a#a 13440 0.215 14592 0.249 15872 0.225
16533 0 1 0 0 1 0 0 0 0 y:b 16640 0.226 16896 0.225 17152 0.211

file: 5140
11514 1 1 0 0 1 0 0 1 vi:k 11776 0.235 12288 0.221 12928 0.211
13251 0 1 0 0 0 0 1 1 ka# 13696 0.233 14848 0.210 15872 0.210
16473 0 1 0 0 0 0 0 0 1 IC 16640 0.221 17024 0.218 17408 0.219

42810 1 0 1 1 0 0 0 0 ? 42752 0.223 42752 0.223 42752 0.223
    
```

Abb. 4: Beispiel Datenkodierung aus [Pes97]

Eine kodierte Silbe belegt hierbei je eine Zeile. Die erste Spalte markiert den Beginn als Sample der Zeitfunktion, die folgenden acht Spalten zeigen den Attributvektor am Netzeingang (ph..phrase boundaries, wb..word boundaries, etc. ..) und in der nächsten Spalte ist die SAMPA-Umschrift für die Silbe zu finden. Die folgenden drei Wertepaare setzen sich jeweils aus Zeitpunkt der f0-Bestimmung (Sample) und dem absoluten f0-Wert zusammen.

Der beschriebene neuronale Ansatz wurde für den Stuttgarter Neuronale Netz-Simulator {SNNS95} aufbereitet und exemplarisch mit den bereits erwähnten Sätzen zur Zugauskunft aus PhonDat trainiert. Es wurden vorerst nur ein Sprecher ausgewählt und die 63 gelabelten Sätze in Training-, Test- und Validation-Set (40, 3, 20) unterteilt. Zur Vermeidung einer Überadaptation auf dieses an sich zu kleine Training-Set wurde das Abbruchkriterium der Lernphase im Test-Set ausgewertet. Eine Plausibilitätsprüfung der praktischen Ergebnisse und erste perzeptive Versuche zeigten die prinzipielle Eignung des neuronalen Verfahrens. Insgesamt wird durch das Verfahren primär die Makrointonation über den Sätzen abgebildet, weniger die segmentelle Mikrointonation. Abbildung 5 vergleicht die am Original gemessene und stilisierte f0-Kontur, welche z.B. im Training genutzt wird, mit der f0-Netzausgabekontur während der Kannphase.

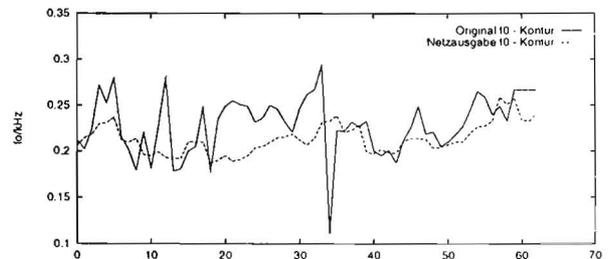


Abb. 5: Original-f0-Kontur versus NN-f0-Kontur für Validation-Beispiel

Die statistische bzw. graphische Bewertung der Ergebnisse - ansatzweise in [Pes97] ausgeführt - ist jedoch proble-

matisch. Die Autoren führten deshalb einen Perzeptions-test durch, dessen Auswertung in die weiteren Arbeiten einfließen soll. Abschließend stellt sich die Frage, welches Verhältnis zwischen neuronaler und externer, symbolischer Verarbeitung sinnvoll ist, oder ob sich ggf. ein hybrider Ansatz eignet.

5 Hörtest & Ergebnisse

5.1 Motivation

Die Skepsis gegenüber neuronalen Intonationsansätzen ist z.B. mit folgenden Fragen verbunden: Wie kann die semantische Korrektheit von Äußerungen bei auftretenden "Fehlakzentuierungen" objektiv bewertet werden? Ist es sinnvoll, ein Distanzmaß zwischen Netzausgabe und realer Lernvorgabe zu definieren, wenn die in der Kannphase gewünschte Ausgabe relativ unscharf ist (z.B. Sprecherintention, individuelle Charakteristika, etc.)? Welche perzeptiven Toleranzen in der f0-Kontur und damit Stilisierungen sind möglich? (Siehe hierzu auch [Adr93].)

Diese Sichtweise ist verständlich, wird aber dem Vergleich mit (ausschließlich) symbolischen Verfahren nicht vollständig gerecht: Natürliche Intonationsdaten werden dort nicht zwingend benötigt sondern liegen als abstraktes Expertenwissen vor. Im Sinne humaner Performanz sind die zuvor aufgeworfenen Fragen für symbolische Verfahren aber genauso relevant.

Ziel des durchgeführten Hörtests war es, einer in Abschnitt "2 Netze kontra Regeln?" angedeuteten und subjektiv motivierten, Detaildiskussion zur Performanz der Intonationsmodelle vorzubeugen und gewisse Vorbehalte (bei Randbedingungen der Sprachsynthese) praktisch zu widerlegen. Der Test bestand aus den drei Teilen A, B und C.

5.2 Beschreibung

Teil A (Satzvergleich) 10 Paare synthetischer Sätze wurden ungeordnet präsentiert. Jeweils ein Satz wurde mittels linearer, regelbasierter Intonation (RG-Into) - der andere mit der neuronalen Intonation (NN-Into) generiert. Frage: "Welcher Satz gefiel Ihnen melodisch besser?". Die 10 verschiedenen Sätze entsprechen den kanonischen Texten aus dem verwendeten NN-Datenmaterial (3x Training-Set, 3x Test-Set, 4x Validation-Set), enthielten je 5 Frage- und 5 Aussagesätze und waren aus 2 kurzen, 6 mittleren und 2 mehrphrasigen Sätzen zusammengesetzt. Die dargebotenen Satzpaar-Realisierungen sind zufällig ausgewählt und repräsentieren einen Querschnitt der trainierten Zugauskunftsdomäne. Drei Satzpaare (1x Training, 1x Test, 1x Validation) wurden wiederholt und dazu "gemischt", um die Urteilszuverlässigkeit zu messen (Reproduzierbarkeit). Exemplarisch wurde auch die stilisierte Original-f0-Kontur (OR-Into) synthetisiert und für einen Satz in Kontrast zur RG-Into und für einen anderen Satz in Kontrast zur NN-Into dargeboten. Insgesamt hatte der Hörer 15 ungeordnete Satzpaare

spontan zu beurteilen, die folgender (geordneter) Übersicht entsprechen:

1. Trn Geht heute noch ein Zug nach Hannover?
NN vs. RG 1b) Wdh
2. Trn Ich möchte heute abend nach Köln fahren.
NN vs. RG
3. Trn Ah ja, am Vormittag, so um 8 Uhr
vielleicht.
NN vs. RG
4. Tst Ich muß über Hannover nach Hamburg
fahren.
NN vs. RG 4b) Wdh
5. Tst Wie kann ich von Ulm nach Düsseldorf
kommen?
NN vs. RG
6. Tst Kann man heute noch von Ulm nach
Koblenz kommen?
NN vs. RG
7. Val Ich muß nach München.
NN vs. RG 7b) Wdh
8. Val Aber sicher.
NN vs. RG
9. Val Kann ich am Samstagabend nach 22 Uhr
noch von Frankfurt nach München
kommen?
NN vs. RG 9b) OR vs. RG
10. Val Gibt es eine Zugverbindung heute abend
nach Frankfurt und wenn ja, auf
welchem Gleis fährt der Zug ab?
NN vs. RG 10b) OR vs. NN

Teil B (Textvergleich) In diesem Teil sollten zwei kurze Texte melodisch miteinander verglichen werden: "Guten Tag, ich möchte eine Auskunft. Und zwar eine Verbindung von Regensburg nach Hamburg. Mir fehlt noch die Verbindung für Samstag früh - morgens zwischen 8 und 9 - also nach 8.". Im Gegensatz zu Teil A wurde hier die Akustische Synthese aus dem Verbmobil-Teilprojekt 4 [VeM95] und die entsprechende RG-Into verwendet, welche allerdings in etwa dem RG-Modell aus Abschnitt 3 entsprechen dürfte. Für den zweiten Text wurde das NN-Into-Verfahren eingesetzt. Die Reihenfolge RG-NN-Text wurde über verschiedene Hörer variiert, um positionelle Vorzugsentscheidungen auszuschließen. Die Frage lautete: "Welches Textbeispiel bevorzugen sie melodisch?".

Teil C (Vergleich verschiedener Akustischer Synthesen) Dem Hörer wurden vier verschiedene Konkatenations-Sprachsynthesen im Paarvergleich präsentiert. Für die Evaluation des NN-Modells besitzt dieser Teil keine große Bedeutung.

5.3 Ergebnisse

40 Hörer (31x männlich, 9x weiblich, mittleres Alter: 36.1) nahmen an dem Test teil. Diese können in folgende Gruppen eingeteilt werden:

- Gruppe GI: 19 (15x m, 4x w), Hörer mit Sprachkommunikationserfahrung
- Gruppe GII: 13 (11x m, 2x w), Hörer mit Akustikkenntnissen
- Gruppe GIII: 8 (5x m, 3x w), Hörer ohne Vorkenntnisse

In Teil A entfielen von insgesamt 400 Ersturteilen (ohne wiederholte Präsentation der Satzpaare) 284 auf die NN-Intonation, das entspricht einer Pro-Quote von 71.0%. Die Entscheidungen sind über GI..GIII und über die Sätze recht gleichmäßig verteilt, wie in Abbildung 6 dargestellt.

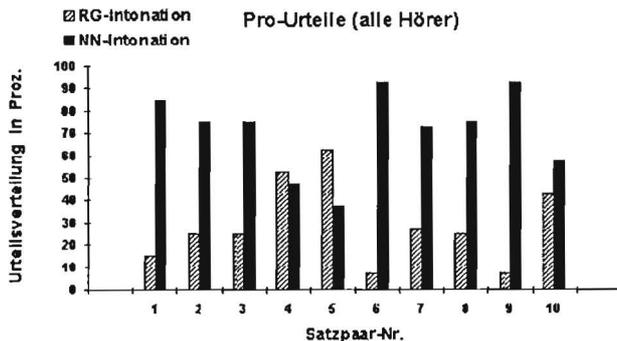


Abb. 6: Pro-Urteile im Satzpaarvergleich (alle Hörer)

Die zwei NN-Into-Verlierer - Satz 4 (Trn) und Satz 5 (Tst) - erhalten immerhin 47.5% bzw. 37.5% der Urteile. Die recht signifikante Pro-NN-Entscheidung überrascht insofern, als daß eine melodische Monotonie der RG-Intonation theoretisch vorrangig für Teil B (Textvergleich) zu erwarten war und GI-Expertenhörer verbal die RG-Intonation als korrekt empfanden, dagegen der NN-Intonation (bei Kenntnis der Quelle) Fehlakzentuierungen zuordneten.

Ein weiterer verbaler Kritikpunkt einiger GI-Hörer war der in der NN-Intonation zu schwache "Final Fall" bei Aussagesätzen. Auch diese Meinung wurde durch den Test nicht bestätigt. Die NN-Aussage-Sätze schneiden mit Pro 69.0% nur unwesentlich schlechter als die NN-Fragesätze (73.0%) ab.

Weitere Aussagen für Teil A lassen sich ableiten, wenn die Urteilsreproduzierbarkeit berücksichtigt wird. Unter den 40 Hörern gibt es 14 RpIII-Hörer (davon 10 aus GI), die bei den drei wiederholt präsentierten Satzpaaren eine identische Zweitentscheidung trafen. Die RpIII-Hörer favorisierten die NN-Varianten sogar mit 79.3% über alle 140 Ersturteile. Die 2 NN-Verlierer werden allerdings von diesem Hörerkreis mit Pro 42.9% bzw. Pro 50.0% bestätigt. Abbildung 7 zeigt die Urteilsverteilung der "zuverlässigen" RpIII-Hörer über die Satzpaare analog zu Abbildung 6.

Teil A sah primär den Test der alternativen RG/NN-Modelle für die Sprachsynthese vor, es fehlt ein ausführlicher Test gegen die stilisierten Original-f0-Konturen (OR). Die beiden präsentierten Satzpaare sollen deshalb

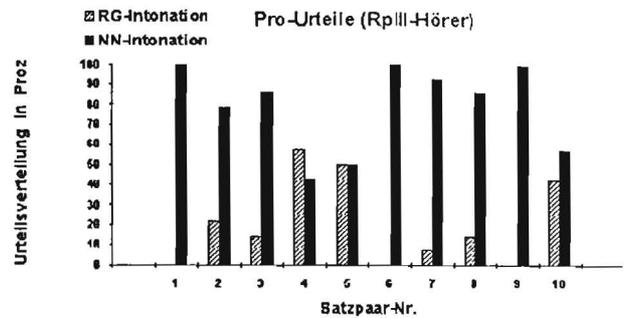


Abb. 7: Pro-Urteile im Satzpaarvergleich (nur RpIII)

lediglich als Anhaltspunkt dienen: Satzpaar 9 ergab für die Präsentation OR versus RG eine klare Stimmverteilung von 92.9% zu 7.1% (exakt das gleiche Ergebnis wie NN vs. RG). Bei Satzpaar 10 bevorzugten nur 57.5% der Hörer OR versus NN (exakt der gleiche Anteil an Hörern wie NN versus RG). Eine mögliche Interpretation dieses exemplarischen Tests: Die "natürliche", stärker variierende OR-Kontur wird bei Sprachsynthese nur bedingt "natürlicher" als die NN-Kontur eingeschätzt. Bei Verwendung resynthetisierter Äußerungen von Original-Zeitfunktionen ergäbe sich eine signifikante OR-Bevorzugung, wie andere Tests der Autoren zeigen. In Teil B entfallen 85.0% der Pro-Stimmen auf den NN-Text - innerhalb der Klasse der RpIII-Hörer aus Teil A sogar 92.9%. Unter Berücksichtigung der veränderten Akustischen Synthese (anderer Sprecher) und veränderten RG-Intonation kann dies als Beleg für die Übertragbarkeit der Ergebnisse aus Teil A gewertet werden. Die starke Bevorzugung des NN-Textes ist vermutlich darauf zurückzuführen, daß die dem RG-Ansatz innewohnende melodische Monotonie bei Satzsequenzen stärker auffällt. Die Signifikanz der Urteile steht in Kontrast zu der häufig in den Versuchsprotokollen verbalisierten Meinung: "Ich konnte kaum Unterschiede feststellen.". Entgegengesetzt wurde in Teil C (Paarvergleich verschiedener Akustischer Synthesen) mehrfach verbal protokolliert: "Qualitätsunterschiede so groß, daß Urteil stets eindeutig...". Die Stimmverteilung z.B. zwischen erst- und zweitplatzierter Synthese hingegen war nicht allzu signifikant (62.5% zu 37.5%). Diese Subjektivität der Urteile in Teil C stützt die signifikanten Urteile aus Teil A bzw. B, wo genau diese erst- bzw. die zweitplatzierte Synthese eingesetzt wurde.

Die Gesamturteile für Teil A und B sind nochmals in Abbildung 8 dargestellt.

6 Schlußfolgerungen

Die verglichenen regelbasierten bzw. neuronalen Verarbeitungsparadigmen basieren auf der gleichen Eingangsinformation und sind im Realisierungsaufwand ähnlich komplex. Die perzeptive Akzeptanz der neuronal modellierten Beispiele ist erstaunlich hoch und widerspricht t.w. der Meinung von Expertenhörern bei Kenntnis des jeweils verwendeten Modells. Der Hörtest unterstreicht

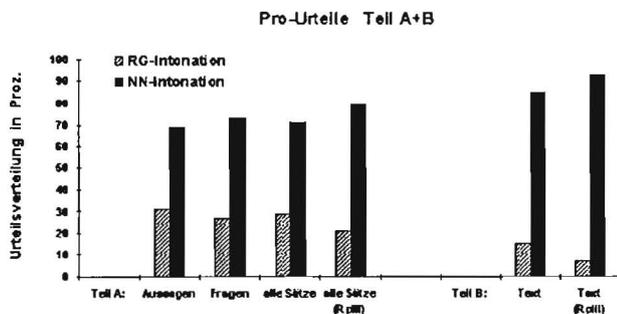


Abb. 8: Vergleich der Gesamt-Pro-Urteile

die Alternative neuronaler Ansätze in der Sprachsynthese. Aufgrund der speziellen Domäne, des geringen Datenumfangs und der relativ geringen Probandenzahl ist eine gezielte Evaluation nicht möglich. Zudem schließt die, durch die Autoren favorisierte, neuronale Strategie, explizite Wissensquellen (z.B. Akzentinformation) und einen gewissen Anteil an symbolischer Verarbeitung mit ein. Für weitere Arbeiten ergeben sich folgende Thesen:

- Auswahl der Verarbeitungsstrategie stärker an Nutzerurteil und praktischen Bedürfnissen orientieren. Expertenurteil als Anregung.
- Beurteilung der Gesamtleistung anstelle der Einzelsysteme.
- Das geeignete Verhältnis verschiedener Methoden in hybriden Ansätzen ist auszutesten.
- Insgesamt mehr Mut zum Ausprobieren!

Die Autoren vervollständigen derzeit das vorgestellte Verfahren zur f₀-Konturplanung, verallgemeinern den Ansatz für eine entsprechende Rhythmussteuerung und planen eine Integration in das komplette Dresdner Text-to-Speech System, wobei dem professionellen Entwickler bzw. Nutzer die Auswahl der Module vorbehalten sein wird (Baukasten-Prinzip).

Literatur

- [Adr93] Adriaens, L.M.H.: Ein Modell deutscher Intonation. Dissertation, TU Eindhoven, 1993
- [Fuj88] Fujisaki, H.: A note on the physiological basis for the phrase and accent components in the voice fundamental frequency contour. In: Osamu Fujimura (Ed.): Vocal physiology: voice production, mechanisms and functions, pages 347-355, Raven Press, New York, 1988
- [Hir96] Hirschfeld, D.: Das Text-to-Speech System. Algorithmen, Programmbeschreibungen, Dateninventare. Forschungsbericht, TU Dresden, 1996
- [Jok95] Jokisch, O.; Kordon, U.: Generierung von Grundfrequenzverläufen in einem Sprachsynthesystem mit neuronalen Netzen. Studententexte zur Sprachkommunikation, Heft 12, S. 113-119, 1995

- [Kor96] Kordon, U.: Ein tragbares Sprachsynthesystem. Spektrum der Wissenschaft, 12: S.113, 1996
- [Pho93] PHONDAT Sprachdatenkorpora auf CD-ROM, Informationsheft. Institut für Phonetik und Sprachliche Kommunikation, München, 1993
- [Pes97] Pescheck, M.: kNN-Ansätze zur Prosodiegenerierung in Text-to-Speech Systemen. Studienarbeit, TU Dresden, 1997
- [Schn94] Schnelle, F.: Einsatz eines prosodischen Modells für einen eingeschränkten Diskursbereich zur Verbesserung der Sprachqualität des Synthetisators VOICE. Diplomarbeit, TU Dresden, 1994
- [SNNS95] SNNS - Stuttgart Neural Network Simulator. User Manual, Version 4.1, Universität Stuttgart, 1995
- [Tra92] Traber, C.: F₀ generation with a database of natural F₀ patterns and with a neural network. In: Talking Machines: Theorie, Models, and Designs, pages 287-304, North-Holland, 1992
- [Tra96] Datengesteuerte Prosodiegenerierung mittels automatischer Lernverfahren. Fortschritte der Akustik - DAGA 96, S. 86-89, 1996
- [VeM95] Karger, R.; Wahlster, W.: Verbmobil Handbuch. Version 3. DFKI GmbH Saarbrücken, 1995

Lokaler vs. Parallel Distribuiertes Konnektionismus in der Modellierung der Sprachproduktion

Marc Kupietz

Fakultät für Linguistik und Literaturwissenschaft

Universität Bielefeld

Postfach 10 01 31, D-33501 Bielefeld

email: kupietz@lili.uni-bielefeld.de

30. Juni 1997

1 Einleitung

Die psycholinguistische Sprachproduktionsforschung auf subkonzeptueller Ebene unterscheidet sich von anderen Bereichen psychologisch orientierter Kognitionsforschung dadurch, daß gerade lokal konnektionistische Ansätze sich in der Modellierung mehr oder weniger durchgesetzt haben. Eine spezifische Ursache für die heute geringe Wichtigkeit rein symbolischer Ansätze ist, daß sich im Gegensatz z.B. zum Sprachverstehen die normale Produktion von Sprache auf lexikalischer und phonologischer Ebene sehr leicht simulieren läßt. So ist es trivial aus einer internen symbolischen Repräsentation von „Your feet are too big.“ die Symbolfolge „Your feet are too big.“ zu generieren (s. Dennett 1991, S. 234f.). Da also „Ausgabeadäquatheit“ als Qualitätskriterium für ein psychologisches Modell der Sprachproduktion nicht ausreicht, müssen andere Domänen von Vergleichsdaten genutzt werden, um die Freiheitsgrade möglicher, alternativer Theorien zu reduzieren. Für die Sprachproduktion bietet sich hier insbesondere an, das bereits vorhandene Wissen über die Systematik von Versprechern auszunutzen und Modelle zu konstruieren, die sowohl normale als auch in diesem Sinne gestörte Produktion erklären und die damit durch die Versprecherdaten zusätzlich restringiert sind. Innerhalb des klassischen regelbasierten Paradigmas scheinen solche Modelle jedoch nicht sinnvoll, da zur Erzeugung von Versprechern letztlich Regeln eingeführt werden müßten, die nichts anderes zu leisten haben, als gelegentlich die Produktion von Fehlern zu erzwingen.

Ein weiterer Nachteil rein symbolischer Ansätze, der gerade in der Untersuchung der Sprachproduktion offensichtlich wird (siehe obiges Beispiel), besteht darin, daß Theorien, die *a priori* die Existenz von symbolischen Repräsentationen oder gar die Existenz einer *Language of Thought* annehmen, verdächtig sind, letztlich nur zirkuläre Erklärungen liefern zu können. In

diesem Zusammenhang scheint es verwunderlich, daß auch verteilt konnektionistische Netze, die in bezug auf repräsentationelle Vorurteilsfreiheit noch weiter gehen als ihre lokalen Konkurrenten, heute nur eine kleine Rolle auf dem Forschungsgebiet spielen.

Ich werde im folgenden den bisher einzigen PDP-Ansatz zur Modellierung von Aspekten der Sprachproduktion (Dell, Juliano & Govindjee 1993) kurz vorstellen, um dann den potentiellen Erklärungswert von auf diesem Ansatz basierenden distribuierten Sprachproduktionsmodellen im Vergleich zu lokalen zu diskutieren.

2 „Frame-Constraints“

Eine Standardannahme der verschiedenen Theorien der Sprachproduktion ist, daß die Generierung einer Äußerung einen Prozeß erfordert, der getrennt repräsentierte linguistische Strukturen und linguistische Inhalte zusammenfügt. Auf der syntaktischen Ebene wird z.B. angenommen, daß zunächst eine Phrasenstruktur des Satzes anhand semantischer Repräsentationen aufgebaut wird. Dieses Gerüst („Frame“) enthält freie Positionen — sogenannte „Slots“ — für Lexeme bestimmter grammatischer Kategorien. Die geeigneten Lexeme werden dann aus dem mentalen Lexikon abgerufen und in die Slots gefüllt. Es gibt also eine Trennung zwischen Struktur (dem Frame) und Inhalt (den Lexemen).

Ein analoger Frame-Slot-Mechanismus wird sowohl in konnektionistischen (Berg 1988; Dell 1988; MacKay 1987; Schade 1992; Stemberger 1985) wie auch in eher symbolischen Ansätzen (z.B. Meyer 1991; Shattuck-Hufnagel 1979) für die phonologische Realisierung von Lexemen angenommen: Ein phonologischer Frame wird zunächst entweder nach den phonotaktischen Regeln der jeweiligen Sprache zusammengesetzt oder komplett aus einem Inventar abgerufen. Danach werden phonologische Segmente in den Frame eingefügt. Diese Theorien postulieren also eine Trennung zwischen der phonologischen Struktur und dem phonologischen Inhalt eines Wortes.

Die Postulation solcher Frame-Slot-Mechanismen stützt sich im wesentlichen auf Versprecherdaten, genauer gesagt auf sogenannte phonologische Frame-Constraints, die eine generalisierende Interpretation von Versprecher-Phänomenen darstellen (s. Tab. 1).

3 Ein PDP-Ansatz zu Erklärung von Frame-Constraints

Dell et al. (1993) haben demonstriert, daß viele der psychologischen Daten, von denen angenommen wurde, daß sie eine phonologische Struktur-Inhalt-Trennung unterstützen, durch einen Mechanismus erklärt werden können, der ohne eine solche explizite Trennung auskommt. Ihr einfach rekurrentes Netz („SRN“, vgl. Elman 1990), das trainiert wurde, arbiträre Lexemrepräsentationen auf Sequenzen entsprechender feature-basierter Phonemrepräsentationen abzubilden, erzeugt, wenn es durch Rauschen gestört wird, ein Fehlerprofil, das in bezug auf die Frame-Constraints weitgehend den in der Versprecherliteratur berichteten Standards entspricht. Das ursprüngliche Modell hatte allerdings die erhebliche Einschränkung, keine Bewegungsfehler (Antizipationen, Perseverationen und Metathesen), die 69% aller Versprecher ausmachen (Garnham, Shillcock, Brown, Mill & Cutler 1981; Dell et al. 1993, S. 184), produzieren zu können.

Effekt	Erläuterung	Standardinterpretation
<i>Phonotaktische Regularität</i>	Phonologische Versprecher erzeugen fast immer in der jeweiligen Sprache vorkommende Lautsequenzen (siehe z.B. Wells 1951).	Phonotaktisch irreguläre Sequenzen werden durch die Anwendung von Frames und Regeln ausgeschlossen.
<i>Kategorie-Erhaltung</i>	Fehler- und Ziel-Segmente gehören praktisch immer derselben Konsonant-Vokal-Kategorie an.	V-Slots und C-Slots akzeptieren nur entsprechende V- bzw. C-Segmente
<i>Silben-Konstituenten-Effekt</i>	Wenn aneinander angrenzende Vokale und Konsonanten in einem Versprecher gemeinsam ersetzt werden, handelt es sich im Englischen wahrscheinlicher um eine VC- als um eine CV-Kombination (vgl. Nootboom 1969; Shattuck-Hufnagel 1983; Stemberger 1983).	CVC-Silbenframes weisen eine strukturelle Trennung zwischen Onset und Rhyme auf.
<i>Initialeffekt</i>	Konsonanten am Wort- und Silbenanfang fallen häufiger Versprechern zum Opfer als Konsonanten an anderen Positionen.	Aufgrund der Onset-Rhyme-Struktur von CVC-Silben-Frames, sind Onset Konsonanten leichter abtrennbar.

Tabelle 1: Frame-Constraints bei Fehlern ohne Segmentbewegungen

Für die gute Reproduktion von Frame-Constraints führen Dell et al. zwei Hauptgründe an. Zum einen ist es sehr wahrscheinlich, daß wenn das Modell ein falsches Segment ausgibt, dieses dem Ziel-Segment phonetisch *ähnlich* ist, da als Ausgabe-Repräsentation phonetische Features gewählt wurden und damit die Ähnlichkeit zweier Phoneme durch den euklidischen Abstand ihrer Repräsentationen modelliert ist. Der größere mittlere Abstand zwischen Vokalen und Konsonanten als zwischen Vokalen und Konsonanten untereinander erklärt auch den Kategorie-Erhaltungs-Effekt. Der zweite wichtige Faktor ist die Tendenz des Modells, „gewohnte“ Sequenzen zu produzieren. Wenn z.B. zuletzt ein Stop-Konsonant produziert wurde, befindet sich in der externen Feedback-Schicht dessen Kopie. In der Trainingsphase sollte das Netz „gelernt“ haben, daß auf einen Stop nie ein weiterer Stop folgt. Der SRN-Mechanismus extrahiert gewissermaßen die stochastischen Haupteigenschaften des Trainingsmaterials. So konnte in verschiedenen Simulationsstudien auch beobachtet werden, daß frequentes Vokabular eine adäquater Performanz bedingte als infrequentes Vokabular, welches die Lautstruktur der englischen Sprache weniger gut reflektiert. Ein deutlicherer Silben-Konstituenten-Effekt in Bedingungen mit frequenten Vokabular liegt z.B. darin begründet, daß im frequenten Vokabular, wie auch im Englischen im allgemeinen, VC-Sequenzen redundanter sind als CV-Sequenzen. Das heißt, es gibt mehr Wörter (bzw. Token), die die gleiche VC-Sequenz beinhalten als Wörter, die die gleiche CV-Sequenz beinhalten. Aus diesem Grund ist die Wahrscheinlichkeit eines Fehlers beim Übergang von C auf V größer als beim Übergang von V auf C. Die Reproduktion des Initialeffekts kann damit begründet werden, daß zu Beginn der Produktion eines Wortes das Netzwerk weniger Informationen hat, da die Kontext-Schichten noch auf neutrale Werte initialisiert sind.

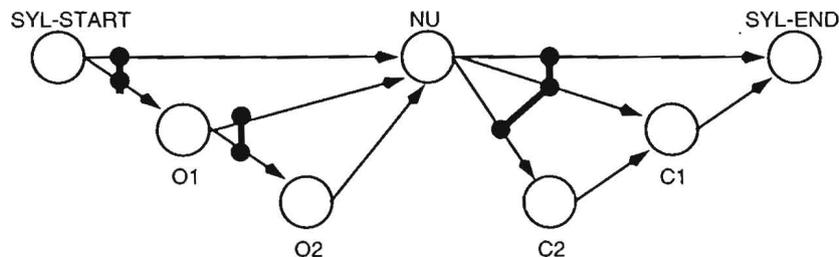


Abbildung 1: Ein Kontrollknotennetzwerk für die Produktion von Silben (Eikmeyer et al. 1995, S. 22). Kreise repräsentieren Kontrollknoten, schwarze Punkte Ventilknöten, Pfeile exzitatorische Verbindungen und fette Linien gegenseitige Inhibition.

4 Ein Vergleich mit einem lokal konnektionistischen Ansatz

In Schades (1992) lokalem konnektionistischen Modell wird die Sequentialisierung von Segmenten durch einen *Kontrollknotennetz-Mechanismus* (Eikmeyer, Schade & Kupietz 1995; Eikmeyer et al. 1995) geleistet. Kontroll-Netze, wie in Abb. 1 dargestellt, bestehen aus einer Anzahl von *Kontrollknoten*, die im Falle des Silben-Netzes die phonologischen Kategorien *Onset1*, *Onset2*, *Nucleus*, *Coda1* und *Coda2* repräsentieren. Jeder dieser Knoten ist mit gerichteten exzitatorischen Leitungen zu Phonemen seiner Kategorie verbunden. Während der Produktion einer Silbe ist immer genau einer der Kontrollknoten aktiv. Nach einer gewissen Zeit wird das Phonem mit der höchsten Aktivierung zur Produktion ausgewählt. Dies ist wahrscheinlich das richtige Phonem, da nur dies sowohl von dem Kontrollknoten als auch vom Knoten der zu äußernden Silbe aktiviert wird. Nach der Selektion leitet der Kontrollknoten seine Aktivierung an einen Nachfolger weiter. Welcher dies ist, hängt von den sogenannten „Gating“- oder *Ventilknoten* ab. Die Ventilknöten beziehen ihre Aktivierung aus der Phonemschicht, so wird z.B. das Ventil in Richtung *Onset2* von allen Phonemen aktiviert, die im Prinzip innerhalb eines Onsetclusters an zweiter Stelle stehen können ($/l/O_2, /r/O_2$)¹. Außerdem sorgen unterschiedliche Ruhenniveaus und eine starke laterale Inhibition zwischen den Ventilknöten dafür, daß in der Regel der richtige Ventilknöten maximal aktiviert ist und damit der richtige Kontrollknoten ausgewählt wird.

Die vier Frame-Constraints sind innerhalb des Modells von Schade folgendermaßen zu erklären:

1. *Phonotaktische Regularität* ist zum Teil im Kontrollknoten-Mechanismus präkodiert. Silben ohne Nucleus sind aufgrund der Netzstruktur „a-priori“ ausgeschlossen; illegale Konsonanten-Kombinationen werden teilweise durch die getrennte Repräsentation von Konsonanten vermieden. Irreguläre Onset-Cluster, wie z.B. $/dl/$, sind im Prinzip möglich, aber unwahrscheinlich, da daß im Netzwerk vorhandene interlaterale Feedback sowohl ähnliche Fehler-Wörter/-Silben wie ähnliche Fehler-Phoneme begünstigt. Es ist also wahrscheinlicher, daß ein $/kl/$ -Onset durch $/gl/$ als durch $/dl/$ ersetzt wird, da $/dl/$ im Gegensatz zu $/gl/$ („lexical Bias“) kein Feedback von der Lexem- bzw. Silben-Ebene bekommt und

¹Das Modell trennt zwischen Konsonanten an verschiedenen Silbenpositionen. Für das Phonem $/r/$ gibt es z.B. im Netz die Knoten $/r/O_1, /r/O_2, /r/C_1$ und $/r/C_2$.

zweitens /g/ wegen seiner phonetisch größeren Ähnlichkeit zu /k/ mehr Feedback von der Ebene der phonetischen Features bekommt.

2. *Kategorie-Erhaltung* ist analog zur phonotaktischen Regularität zu erklären.
3. Der *Silbenkonstituenz-Effekt* ist ähnlich wie bei Dell et al. auf die stochastischen Eigenschaften der Struktur des Vokabulars zurückzuführen. In lokal konnektionistischen Modellen sind Wortfrequenzen meist durch Ruhewerte von Lemma-Knoten modelliert (vgl. Dell 1990, S. 331ff). Häufig zusammen auftretende Silbenkonstituenten geben einander proportional mehr Feedback. Daher ist der Übergang zwischen ihnen weniger fehleranfällig.
4. Der *Initialeffekt* ist ähnlich wie bei Dells Modell Ausdruck einer geringeren Präsenz von „Information“ im Netz. Der Mangel betrifft allerdings weniger Sequenzinformationen, sondern eine noch nicht völlig ausgebreitete und „gesetzte“ Aktivierung im Inhaltsnetz.

Zusammenfassend läßt sich feststellen, daß auch bei dem in Eikmeyer et al. (1995) beschriebenen Ansatz die Reproduktion von Frame-Constraints im wesentlichen nicht „a priori“-Modellannahmen reflektiert, sondern auf die Sensitivität des Netzwerks für das zugrundeliegende Vokabular zurückzuführen ist. Der wesentliche qualitative Unterschied zwischen den Ansätzen in bezug auf die Erklärung von Frame-Constraints besteht darin, daß erstens in dem lokalen Modell mehr zu motivierende Annahmen über Struktur und Repräsentation des Vokabulars gemacht werden und es zweitens dort keinen uniformen Mechanismus für Paradigmatik und Syntagmatik gibt. Allerdings läßt sich anhand dieser Unterschiede aufgrund der extremen Begrenztheit des erklärten Phänomens noch kein qualitativer Vergleich der Ansätze vornehmen, zumal Schades Modell eine wesentlich größere Datendomäne erklärt.

Eine der größten Einschränkungen des Ansatzes von Dell et al. ist wie gesagt, daß nur solche Versprecher erklärt werden, die keine Segmentbewegungen beinhalten. Diese Einschränkung ist besonders kritisch, da diese nicht-kontextuellen, sogenannten „Non-Movement“-Fehler nur etwa 31% aller Fehler ausmachen. Dell et al. geben als möglichen Lösungsansatz eine Sichtweise an, in der kontextuelle Einflüsse ebenso wie andere fehlerverursachende Einflüsse, z.B. mangelnde Aufmerksamkeit, generell als Teil des Rauschens betrachtet werden, der sich auf das Netz auswirkt. Um kontextuelle Einflüsse systematisch zu modellieren, wird vorgeschlagen, das Eingabe-Muster eines Zielwortes mit den Repräsentationen angrenzender Wörter zu kontaminieren (siehe Dell et al. 1993, S. 184). Eigene Simulationsstudien (Kupietz 1996) haben allerdings gezeigt, daß auch so kein adäquater Anteil an Bewegungsfehlern erzielt werden kann.

Abgesehen von den empirischen Befunden sprechen auch theoretische Überlegungen gegen die Erweiterbarkeit des Ansatzes zu einem allgemeinen Sprachproduktionsmodell. Die von Dell et al. vorgeschlagene Kontamination von Inputs ist z.B. schwer zu motivieren, da Antizipations-Fehler auf einen Mechanismus zurückzuführen sein sollten, der flüssiges Sprechen unter normalen Umständen erleichtert (eine ausführliche Diskussion findet sich bei Dell, Burger & Svec 1997). Im Sprachproduktionsmodell von Schade (1992) sind z.B. das Vorwärmen von zu produzierenden Einheiten und Antizipationsfehler auf einen uniformen Mechanismus zurückzuführen. Innerhalb des vorgestellten Ansatzes gibt es keinen Mechanismus, der unter normalen Umständen von der Kontamination der Eingabe-Schicht durch zukünftig zu produzierende Wörter profitieren könnte. Der Erweiterungsvorschlag ist also gewissermaßen „ad hoc“.

Darüber hinaus könnte ein hypothetisch auf diesem Ansatz basierendes „funktionierendes“ Modell zwar Antizipations- und Perseverationsfehler, aber keine Vertauschungsfehler erklären, da innerhalb des rekurrenten Netzes eine Segment-Antizipation die Wahrscheinlichkeit für eine nachfolgende Produktion des Zielsegmentes und damit die Wahrscheinlichkeit für eine Vertauschung nicht beeinflusst. Das dem Modell zugrundeliegende Sequentialisierungsverfahren ist inhärent für Zukunft und Vergangenheit „blind“.

Ein weiterer Nachteil des SRN-Sequentialisierungsmechanismus ist, daß er nicht zuläßt, den zeitlichen Ablauf von Prozessen während, vor und nach der Produktion von Phonemen zu modellieren. Die minimale zeitliche Granularität, also ein Simulationszyklus, entspricht eben genau der Produktion eines Phonems. Dies schränkt die Erklärungsfähigkeit aller auf diesem Mechanismus basierenden Modelle erheblich ein. Außerdem ist dies auch ein Nachteil gegenüber lokal konnektionistischen Modellen, in denen in der Regel eine beliebig feine zeitliche Auflösung möglich ist. In Schades Modell z.B. breitet sich Aktivierung über die Zeit aus, während laterale Inhibition für die Deaktivierung irrelevanter Einheiten sorgt (vgl. Dell 1986; Berg & Schade 1992; Schade & Berg 1992). „Zwingt“ man ein solches Modell zu einer schnellen Abfolge von Selektionen bzw. Produktionen, entstehen spezifische Vorhersagen über das Fehlerprofil unter Zeitdruck gesprochener Äußerungen. Außerdem liefern solche Modelle Erklärungshypothesen über den *Entstehungsprozeß* von Versprechern.

5 Allgemeine Diskussion

Wesentlich für die Konnektionismus–Symbolismus–Debatte war, daß sich aus distribuierten Repräsentationen keine *Symbole* extrahieren lassen. So ist in einem PDP–Netz keine Einheit lokalisierbar, die permanent als für eine bestimmte andere Entität stehend interpretiert werden kann. Innerhalb der theoretischen Diskussion fungierte dies als Argument gegen den Vorwurf, daß Konnektionismus gegenüber Symbolismus keinen wirklich neuen Ansatz darstellte (siehe Fodor & Pylyshyn 1988; Smolensky 1988). Die Tatsache, daß sich lokal konnektionistische Ansätze in dieser Hinsicht nicht in dem Maße von traditionellen unterscheiden, da in ihnen symbolartige Einheiten lokalisierbar sind und sie insofern auf einer Zwischenstufe zwischen symbolischen und PDP-Modellen stehen, ist hauptverantwortlich für ihre geringe Präsenz in der theoretischen Konnektionismus–Literatur. Der Unterschied zwischen den internen Repräsentationen lokaler und distribierter Netze bedarf allerdings einer genaueren Betrachtung. In lokalen konnektionistischen Netzwerken sind zwar die Knoten mit Labels versehen, jedoch spielen diese Labels für die *Funktion* des Netzes keine Rolle. Sie erleichtern lediglich die Erstellung des Modells und die *Interpretation* seines Verhaltens. Weiterhin ist auch eine traditionelle symbolische Interpretation des Zustands eines lokalen Netzes nur *approximativ* möglich, da jeweils mehrere Knoten einer Ebene, bzw. einer Kategorie gleichzeitig aktiv sein können. Eine getreue Interpretation der Phonemschicht eines Netzes zu einem Zeitpunkt t muß also auf der Aktivationsverteilung über alle (aktiven) Knoten dieser Schicht beruhen. In dieser Hinsicht ist auch die Information in einem lokalen Netz *distribuiert*.

Der Unterschied zu Backpropagation–PDP–Netzen besteht also hauptsächlich darin, daß die Netzwerkarchitektur, das heißt die Verbindungen interner Knoten untereinander und zu den Ein- und Ausgabeknoten, von Hand festgelegt werden. Der so erstellte Repräsentationsrahmen

ist ein Teil der durch das Modell aufgestellten Hypothesen. Die durch das Modell gelieferten Erklärungen beruhen dementsprechend auf den postulierten Knoten und den postulierten Verbindungen zwischen diesen. Im Falle von auf Backpropagation basierenden PDP-Modellen ist der Gewinn und die Formulierung psychologischer Erkenntnisse über die Struktur interner Repräsentationen und die Dynamik der ablaufenden Prozesse zunächst schwieriger. Aufgrund der nicht festgelegten Bedeutung einzelner interner Knoten und der stark verteilten Information über Knoten und Leitungen sind Erkenntnisse, über die Feststellung hinaus, daß gewisse psychologische Phänomene sich reproduzieren lassen, nur durch eine Umkehrung des gewöhnlichen explanatorischen Prozesses möglich: Ausgehend von einer Simulation (einer Level 3 Theorie in der Marr-Hierarchie) muß untersucht werden, welche Eigenschaften des Modells für seine Funktion hauptsächlich verantwortlich sind, um so eine Erklärung auf einer höheren Ebene zu ermöglichen. Der Wert solcher durch statistische *post hoc* Analysen, wie Cluster- oder Hauptkomponenten-Analyse gewonnen Erklärungen hängt allerdings direkt von der Validität der Annahmen über Ein- und Ausgabe-Kodierung und der Auswahl des Trainingsmaterials ab. Im Falle des Ansatzes von Dell et al. ist nicht offensichtlich, wie z.B. eine psychologisch valide Ausgabe-Kodierung beschaffen sein müßte. Die aus einer *post hoc* Analyse des vorgeschlagenen Netzes gewonnen Ergebnisse würden nur die Annahmen reflektieren, welche beim Netz-Design insbesondere über die Kodierung der Ausgabeschicht getroffen wurden. Im Vergleich zu lokalen Netzen mit vergleichbaren Ein- und Ausgaberepräsentationen wird in bezug auf den Erkenntnisgewinn somit kein Vorteil erzielt.

Eine Perspektive für PDP-Netze in der Sprachmodellierung besteht z.B. darin, auf die Annahme linguistischer Konstrukte wie „Phonem“ oder „phonetisches Feature“ zu verzichten, indem die Ausgabeschicht von Netzen nicht zur Repräsentation dieser verwendet wird, sondern direkt zur Steuerung eines Artikulationsapparats (s. Browman & Goldstein 1995). Die Analyse selbstorganisierter verteilter Repräsentationen, die statt anhand von phonologischen Symbolen anhand von physischen Parametern des artikulatorischen Systems gelernt wurden, könnte Erklärungen liefern, die einen wirklichen „dimensional shift“ (vgl. Smolensky 1988) beinhalten.

6 Zusammenfassung

Es wurde demonstriert, daß einfach rekurrente Error-Backpropagation-Netze keinen geeigneten Ansatz zur Modellierung des kognitiven Prozesses der Sprachproduktion darstellen, insbesondere da sie keinen psychologisch adäquaten Mechanismus zur Linearisierung von Phonemen zur Verfügung stellen. Weiterhin wurde argumentiert, daß verteilte Netze generell nur dann andere Erklärungen liefern können als lokale, wenn sie fundierte, sub-symbolische Ein- bzw. Ausgaben einbeziehen.

Literatur

- Berg, T. (1988). *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell*. Tübingen: Niemeyer.
- Berg, T. & Schade, U. (1992). The role of inhibition in a spreading-activation model of lan-

- guage production. Part I. The psycholinguistic perspective. *Journal of Psycholinguistic Research*, 21(6), 405–434.
- Browman, C. P. & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. Port & T. van Gelder (Hrsg.), *Mind as Motion: Explorations in the Dynamics of Cognition* (S. 175–194). Cambridge, MA: MIT Press.
- Dell, G. S. (1986). A spreading–activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of prediction from a connectionist model. *Journal of Memory and Language*, 27, 124–142.
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5(4), 313–349.
- Dell, G. S., Burger, L. K. & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, 104, 123–147.
- Dell, G. S., Juliano, C. & Govindjee, A. (1993). Structure and content in language production. *Cognitive Science*, 17, 149–195.
- Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown & Co.
- Eikmeyer, H.-J., Schade, U. & Kupietz, M. (1995). Ein konnektionistisches Modell für die Produktion von Objektbenennungen. *Kognitionswissenschaft*, 5, 108–117.
- Eikmeyer, H.-J., Schade, U., Kupietz, M. & Laubenstein, U. (1995). Connectionist syntax and planning in the production of object specifications. In R. Meyer-Klabunde & C. von Stechow (Hrsg.), *Conceptual and Semantic Knowledge in Language Production. Proceedings of a Workshop of the Special Collaborative Research Program 245 "Language and Situation"* (Vol. 92, S. 18–30). Heidelberg/Mannheim.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Fodor, J. A. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3–71.
- Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D. & Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics*, 19, 805–817.
- Kupietz, M. (1996). *Modellierung der Sprachproduktion: Perspektiven neuerer kognitionswissenschaftlicher Ansätze*. Unveröffentlichte Magisterarbeit, Fakultät für Linguistik und Literaturwissenschaften, Universität Bielefeld.
- MacKay, D. G. (1987). *The Organization of Perception and Action*. New York: Springer.
- Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, 30, 69–89.

- Nooteboom, S. G. (1969). The tongue slips into patterns. In A. J. Sciarone, A. J. van Essen & A. A. van Raad (Hrsg.), *Leyden Studies in Linguistics and Phonetics*. Den Haag: Mouton.
- Schade, U. (1992). *Konnektionismus — Zur Modellierung der Sprachproduktion*. Opladen, Germany: Westdeutscher Verlag.
- Schade, U. & Berg, T. (1992). The role of inhibition in a spreading-activation model of language production. Part II. The simulational perspective. *Journal of Psycholinguistic Research*, 21(6), 435–462.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-order mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Hrsg.), *Sentence Processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In *The Production of Speech* (S. 109–136). New York: Springer Verlag.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1–74.
- Stemberger, J. P. (1983). *Speech Errors and Theoretical Phonology: A Review*. Bloomington, IN: Indiana University Linguistics Club.
- Stemberger, J. P. (1985). An interactive activation model of language production. In A. Ellis (Hrsg.), *Progress in the Psychology of Language* (Vol. 1, S. 143–186). London: Erlbaum.
- Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine*, 3, 9–30.

Heuristische Ergebnisbewertung beim Wortagentenparsing

Andreas Mertens
FernUniversität Hagen
Praktische Informatik VII/Artificial Intelligence
D-58084 Hagen, Germany
email: andreas.mertens@fernuni-hagen.de

1 Einleitung

Ein zentrales Problem bei der maschinellen Sprachverarbeitung besteht darin, daß die verwendeten Verfahren zur syntaktisch-semanticen Analyse eines natürlichsprachlichen Ausdrucks oft nicht nur eine Ergebnisstruktur, sondern eine große Zahl alternativer Ergebnisstrukturen liefern. Dieses Problem besteht besonders bei inkrementellen Verfahren, da sie versuchen, für jeden bereits eingelesenen Teilausdruck eine Ergebnisrepräsentation aufzubauen, obwohl der zu analysierende sprachliche Ausdruck noch nicht vollständig bekannt ist.¹

Um dieses Problem zu lösen, wird ein zweistufiges Analysemodell vorgeschlagen. Auf der ersten Stufe werden wortweise inkrementell alle Interpretationen zu einem Eingabeausdruck erzeugt. Da in diesem Schritt das kombinatorische Potential der zugrundeliegenden Grammatik voll ausgenutzt wird, ist die Anzahl der zurückgelieferten Ergebnisstrukturen dementsprechend hoch. Auf der zweiten Stufe wird ein heuristisches Verfahren zur Auswahl der optimalen Ergebnisstruktur(en) eingesetzt. Dabei handelt es sich um ein symbolisches Verfahren, welches u. a. Kriterien wie Phrasenanzahl, Phrasenkomplettierung und Verschachtelungstiefe natürlichsprachlicher Ausdrücke berücksichtigt.

Ein Vorteil des hier vorgeschlagenen zweistufigen Modells ist, daß es eine klare Trennung zwischen der Repräsentation von sicherem Wissen (Grammatik) und unsicherem Wissen (heuristische Auswahl aus alternativen Strukturen) ermöglicht. Ein Nachteil des Modells ist, daß es ausschließlich symbolische Kriterien einbezieht. Somit lassen sich z. B. bestimmte Anbindungsmehrdeutigkeiten von Präpositionalphrasen nicht eindeutig auflösen, obwohl die syntaktisch-semanticen Analyse der ersten Stufe auf ausführliche Subkategorisierungsinformation bei Verben, Nomen und Präpositionen zurückgreifen kann. Gerade in diesem Bereich könnte die Einbeziehung von statistischen Verfahren bzw. Ergebnissen weiterhelfen. Deshalb ist geplant, das heuristische Verfahren zu einem hybriden Verfahren weiterzuentwickeln, welches bei der heuristischen Disambiguierung neben symbolischen Kriterien zusätzlich statistisch ermittelte Anbindungspräferenzen berücksichtigt. Im Beitrag werden zunächst die grundlegenden Prinzipien des verwendeten Analyseansatzes mit Wortagenten und des heuristischen Verfahrens vorgestellt, um dann eine mögliche Einbeziehung von statistischen Aspekten am Beispiel der PP-Anbindung aufzuzeigen.

¹In den letzten Jahren zeichnet sich ein deutlicher Trend zu inkrementellen Verfahren in der maschinellen Sprachverarbeitung ab (vgl. z. B. [Hau 89], [MC 94], [Han 95]). Zur Diskussion kognitiver Aspekte inkrementeller Verarbeitung vgl. z. B. [Hem 93].

2 Einordnung des Parsingverfahrens

Im Bereich der automatischen Sprachverarbeitung gibt es Analyseansätze, die die Repräsentation grammatischen Wissens in Form von Wortexperten vorschlagen. Dazu gehören u. a. das von Small entwickelte Wortexperten-Parsing, die wortklassengesteuerte funktionelle Analyse von Helbig, das wortorientierte Parsen von Eimermacher und das lexikalisch verteilte Text-Parsing von Hahn. In der Theorie des Wortexperten-Parsing wird die Verarbeitung natürlicher Sprache als verteilter Prozeß von interagierenden Wörtern betrachtet. Jedes Wort stellt einen aktiven Prozeß dar, den sog. Wortexperten, der unter Zuhilfenahme von linguistischem und Weltwissen seine eigene Bedeutung im aktuellen Kontext mit anderen Wortexperten aushandelt (vgl. [Sma 87]). Das von Eimermacher entwickelte wortorientierte Parsen basiert auf der Idee der Wortexperten, unterscheidet allerdings zwischen Wortklassenexperten, die die grammatischen Informationen repräsentieren, und Wortexperten, die die Beziehungen zwischen den einzelnen Wörtern eines Satzes analysieren (vgl. [Eim 88]).

Neuere Ansätze verbinden das Wortexpertenmodell mit einer agentenorientierten Modellierung der Sprachverarbeitungsprozesse. Beim *ParseTalk*-Modell von Hahn, Schacht und Bröker handelt es sich um ein objektorientiertes Parsing-System, in dem das grammatische Wissen vollständig in das Lexikon integriert ist. Jedes Wort eines gerade analysierten Satzes aktiviert einen Wortaktor, der mit anderen bereits initialisierten Wortaktoren kommuniziert, um so die Gesamtbedeutung des Satzes zu konstituieren (vgl. [HSB 94]). Beim Wortagentenparsing besteht die Grundidee darin, daß jedes Wort syntaktische und semantische Erwartungen auslöst, die erst im weiteren Verlauf der Analyse erfüllt werden können. Nach Einbeziehung weiterer Wörter in die Verarbeitung werden die hinzukommenden Informationen daraufhin überprüft, ob sie zu den ausgelösten Erwartungen passen (vgl. [HM 94]). Dieser Ansatz wird zur Zeit zu einem robusten Parsingverfahren mit folgenden Eigenschaften weiterentwickelt (vgl. [Mer 97]):

- Wortagenten verarbeiten wortklassen-, wort- und wortformspezifisches Wissen.
- Die Verarbeitung verläuft wortweise inkrementell von links nach rechts.
- Die Disambiguierung wird durch ein heuristisches Verfahren unterstützt.

3 Parsing mit Wortagenten

Für jedes eingegebene Wort bzw. für jede eingegebene Wortform aktiviert die morphologisch-lexikalische Analyse² unter Zugriff auf die lexikalischen Merkmalstrukturen einen bzw. im Falle von lexikalischen Mehrdeutigkeiten mehrere Wortagenten, die jeweils über wortklassenspezifisches (z. B. die für die Wortklasse des Eingabeworts gültigen Grammatikregeln), wortspezifisches (z. B. semantische Sorten und Merkmale, Subkategorisierungsinformation) und wortformspezifisches Wissen (z. B. morpho-syntaktische Merkmale für Kongruenzüberprüfungen) verfügen.

Jeder aktivierte Wortagent erhält als Eingabe eine Menge von Merkmalstrukturen, die jeweils die alternativen Analyseergebnisse der vorhergehenden Wortagenten repräsentieren. Diese Menge faßt die Erwartungen der bereits analysierten Wörter zusammen (vgl. EXP_i in Abbildung 1). Aufgrund von wortklassen-, wort- und wortformspezifischem Wissen werden die übergebenen Strukturen ausgewertet. Dabei versuchen die Wortagenten die von den Vorgängern bereits eingetragenen Erwartungen zu erfüllen (z. B. Komplettierung einer Nominalphrase, Anbindung einer Konstituente an das Verb) bzw. diese zu neuen Erwartungen zu verfeinern (z. B. Erweiterung einer bereits durch einen Artikel eröffneten Nominalphrase durch ein Adjektiv). Das Ergebnis wird dann an den nachfolgenden Wortagenten weitergegeben.

²Als morphologische Analyse wird eine erweiterte Version der in LISP implementierten Analyse MORPH von Hanrieder eingesetzt (vgl. [Han 91]).

In Abbildung 1 sind die durch die morphologisch–lexikalische Analyse aktivierten Wortagenten und das Weiterreichen der wortweise inkrementell erzeugten Analyseergebnisse EXP_i für die Eingabephase *der junge Schüler* dargestellt. Die Ausgabestruktur EXP_i eines Wortagenten wird jeweils heuristisch ausgewertet (vgl. Abschnitt 4) und die dabei ermittelte optimale Lesart $h(EXP_i)$ wird als Ergebnis der Analyse ausgegeben.³ Dem nachfolgenden Wortagenten werden allerdings alle potentiell möglichen Lesarten übergeben, damit es möglich bleibt, daß bestimmte, zunächst unwahrscheinliche Erwartungen durch nachfolgende Wortagenten erfüllt werden können (z. B. bei erweiterten Partizipialkonstruktionen: [der [junge Schüler]_{np} unterrichtende Lehrer Müller]_{np}).

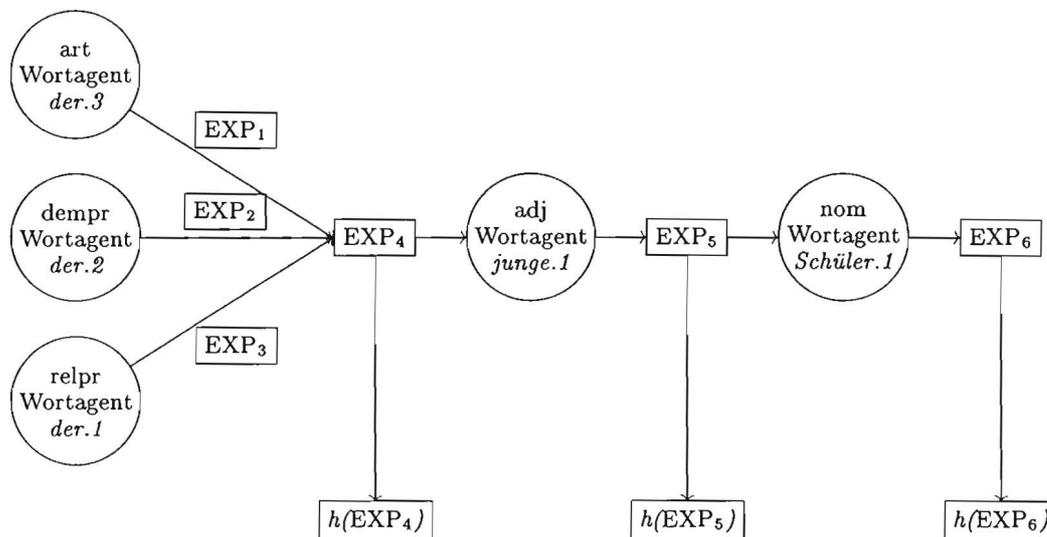


Abbildung 1: Aktivierung von Wortagenten und Weitergabe der Erwartungen

Ist die eingegebene Wortform lexikalisch mehrdeutig, so werden konkurrierende Wortagenten aktiviert, die alternative Interpretationen erzeugen (vgl. EXP₁ bis EXP₃ in Abbildung 1), die dann wieder als Menge von Merkmalstrukturen zusammengefaßt und an den nachfolgenden Wortagenten weitergegeben werden (vgl. EXP₄ in Abbildung 1). Typische Aufgaben der Wortagenten bei der Bearbeitung der Erwartungsstrukturen sind beispielsweise:

- Kongruenzüberprüfungen (bzgl. Kasus, Genus, Numerus etc.) und wortweise inkrementeller Aufbau von syntaktischen Strukturen;
- Markierung der erzeugten (Teil-)Strukturen mit grammatischen Informationen (Erweiterungsregeln);
- Kompatibilitätstests (bzgl. semantischer Sorten und Merkmale) und Aufbau von semantischen Netzen;
- Verarbeitung von alternativen syntaktisch–semantischen Ergebnisstrukturen u. a. m.

4 Heuristische Bewertung der Ergebnisstrukturen

4.1 Ergebnisstrukturen

Schon die Analyse der einfachen Nominalphrase *der junge Schüler* erzeugt bereits sechs verschiedene Interpretationen, für die beispielhaft morpho–syntaktische Merkmalstrukturen in Abbildung 2 dargestellt sind.⁴ Dabei wird unter dem Merkmal CAT nicht nur die zugehörige Ka-

³Evtl. werden auch mehrere alternative Interpretationen ermittelt.

⁴Der zusätzlich aktivierte Wortagent *der.1* für das Relativpronomen trägt aufgrund seiner Stellung im Satz nicht zu den Ergebnisstrukturen bei.

tegorie eingetragen, sondern diese wird zusätzlich dadurch gekennzeichnet, ob es sich um eine eröffnete oder um eine komplettierte Phrase handelt (z. B. np-open, pp-comp). So werden beispielsweise Nominal- und Präpositionalphrasen als "komplettiert" betrachtet, wenn sie durch ein Nomen abgeschlossen sind. Bei Verbalphrasen müssen für die Komplettierung alle obligatorischen Argumente vollständig interpretiert und an das Verb angebunden sein.

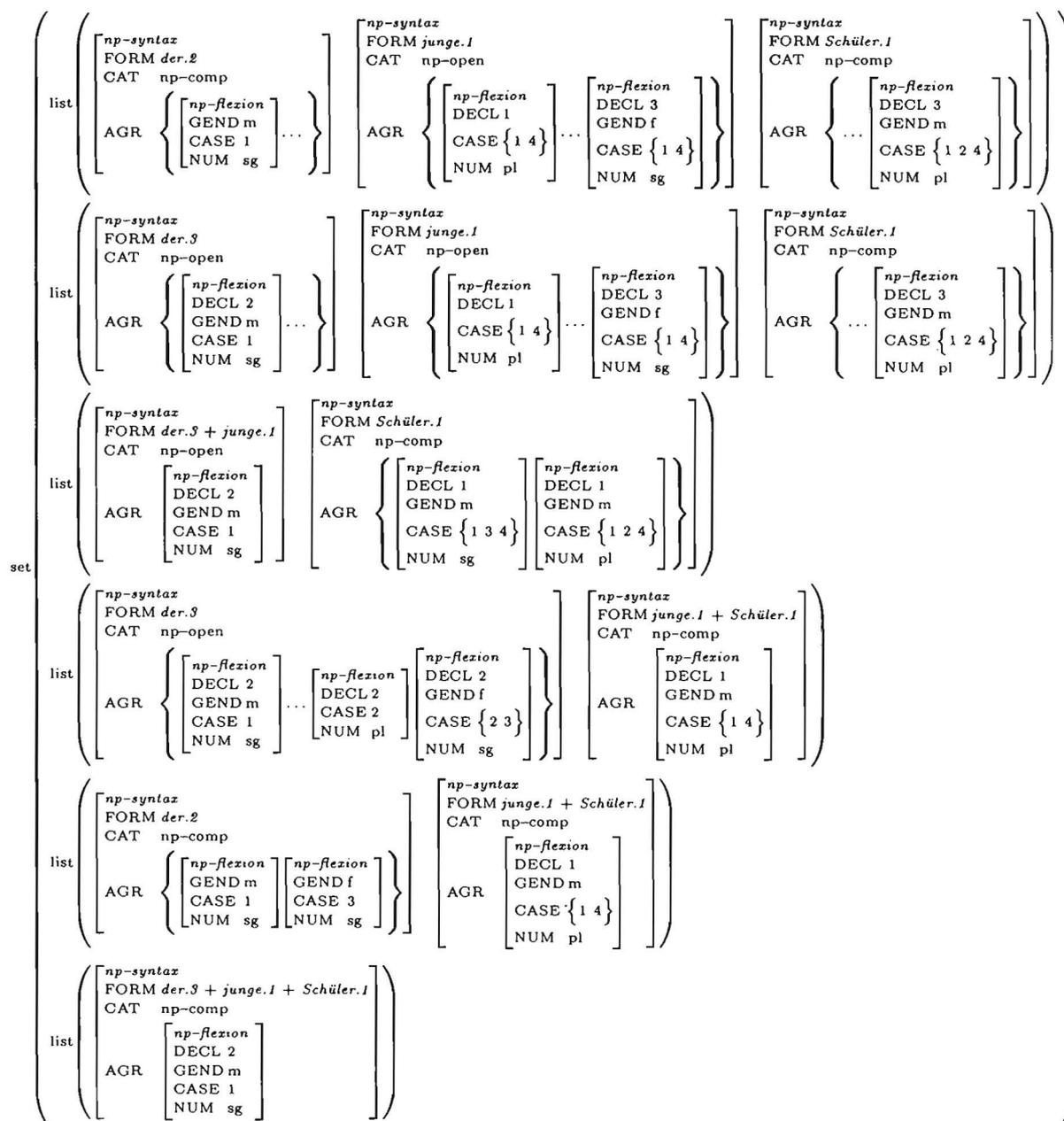


Abbildung 2: Ergebnisstrukturen zu *der junge Schüler*

Im nachfolgenden Abschnitt wird das Basisverfahren zur heuristischen Ergebnisbewertung erläutert. Das heuristische Verfahren wertet die insgesamt erzeugten Ergebnisstrukturen nach symbolischen Kriterien aus. Zur Veranschaulichung des Verfahrens beschränken wir uns auf das in Abbildung 3 angegebene verkürzte Format für Ergebnisstrukturen.⁵

⁵Dabei werden Listenklammern nicht explizit gekennzeichnet. Die Angabe einer Phrase beschränkt sich auf die Auflistung der Merkmalwerte zu CAT und FORM.

$$\text{set} \left(\begin{array}{l} \left(\left(\text{np-comp } der.2 \right) \left(\text{np-open } junge.1 \right) \left(\text{np-comp } Schüler.1 \right) \right) \\ \left(\left(\text{np-open } der.3 \right) \left(\text{np-open } junge.1 \right) \left(\text{np-comp } Schüler.1 \right) \right) \\ \left(\left(\text{np-open } der.3 \right) \left(\text{np-comp } junge.1 \right) \left(\text{np-comp } Schüler.1 \right) \right) \\ \left(\left(\text{np-open } der.3 \right) \left(\text{np-comp } junge.1 \right) \left(\text{np-comp } Schüler.1 \right) \right) \\ \left(\left(\text{np-comp } der.2 \right) \left(\text{np-comp } junge.1 \right) \left(\text{np-comp } Schüler.1 \right) \right) \\ \left(\left(\text{np-comp } der.3 \right) \left(\text{np-comp } junge.1 \right) \left(\text{np-comp } Schüler.1 \right) \right) \end{array} \right)$$

Abbildung 3: Verkürzte Darstellung von Ergebnisstrukturen

4.2 Basisverfahren

Für die Auswahl der wahrscheinlichsten Ergebnisstruktur(en) wird ein heuristisches Basisverfahren eingesetzt, welches die von einem Wortagenten erzeugten Interpretationen bewertet und anschließend daraus die optimale(n) Interpretation(en) auswählt. Das Ziel des Verfahrens ist es, aus den insgesamt erzeugten Strukturen vorrangig diejenigen auszuwählen, die bereits kompletterte Phrasen enthalten.

Zusätzlich wird die Auswahl dahingehend gesteuert, daß die Ergebnisstrukturen möglichst lange Phrasen und somit insgesamt eine geringe Anzahl an Phrasen enthalten. Nachfolgend ist das Verfahren in einer vereinfachten Form angegeben:

- In der Menge M werden alle erzeugten Ergebnisstrukturen zur Eingabephase gesammelt.
- Trage in die Menge M_1 alle Ergebnisstrukturen aus M ein, die nur komplettierte Phrasen enthalten.
- Trage in die Menge M_2 alle restlichen Ergebnisstrukturen aus M ein.⁶
- Wähle aus M_1 die Ergebnisstrukturen aus, die eine minimale Anzahl MIN_1 an komplettierten Phrasen enthalten und entferne alle anderen Ergebnisstrukturen aus M_1 .
- Wähle aus M_2 die Ergebnisstrukturen aus, die eine minimale Anzahl MIN_2 an Phrasen enthalten und entferne alle anderen Ergebnisstrukturen aus M_2 .
- Falls $MIN_1 \leq MIN_2$, wähle M_1 , sonst M_2 .

Als weitere Kriterien, die bei der heuristischen Bewertung zusätzlich eingesetzt, hier aber nicht näher erläutert werden, sind die folgenden zu nennen:

- Bewertung der Verschachtelungstiefe (z. B. bei erweiterten Partizipialkonstruktionen);
- Negativ-Bewertung von Ergebnisstrukturen, die im Vorfeld des Verbs eröffnete Nominal- bzw. Präpositionalphrasen enthalten;
- Bewertung der Kompatibilität von semantischen Merkmalen (z. B. bei der Anbindung von Komplementen an das Verb).

4.3 Analysebeispiel: Erweiterte Partizipialkonstruktionen

Für den Beispielsatz *Der die jungen Schüler unterrichtende Lehrer schreibt.* ergibt sich bei Anwendung des Basisverfahrens auf die jeweiligen inkrementell erzeugten Ergebnisse der einzelnen

⁶Das eigentliche Verfahren unterscheidet hier noch weitere Mengen, z. B. die Menge der Ergebnisstrukturen, die nur eröffnete Phrasen enthalten.

Wortagenten die in Abbildung 4 angegebene Auswahl an Ergebnisstrukturen. Das Beispiel macht deutlich, daß bereits mit einem recht einfachen heuristischen Verfahren die Auswahl der optimalen Ergebnisstruktur(en) gelingt. Wo die Grenzen des verwendeten symbolischen Auswahlverfahrens liegen, soll im folgenden Abschnitt am Beispiel der PP-Anbindung gezeigt werden.

Eingabe: *der*
 Insgesamt erzeugte Strukturen: 2
 Davon ausgewählte Strukturen: 1
 $\text{set}\left(\left(\left(\text{np-comp } \textit{der.2}\right)\right)\right)$

Eingabe: *der die*
 Insgesamt erzeugte Strukturen: 4
 Davon ausgewählte Strukturen: 1
 $\text{set}\left(\left(\left(\text{np-comp } \textit{der.2}\right) \left(\text{np-comp } \textit{die.2}\right)\right)\right)$

Eingabe: *der die jungen*
 Insgesamt erzeugte Strukturen: 6
 Davon ausgewählte Strukturen: 2
 Bemerkung: Das erweiterte Verfahren, welches z. B. noch die Menge der nur eröffneten Phrasen berücksichtigt, reduziert hier die Anzahl auf 1.
 $\text{set}\left(\left(\left(\left(\text{np-comp } \textit{der.2}\right) \left(\text{np-open } \textit{die.3 jungen.1}\right)\right)\right) \left(\left(\text{np-open } \textit{der.3}\right) \left(\text{np-open } \textit{die.3 jungen.1}\right)\right)\right)$

Eingabe: *der die jungen Schüler*
 Insgesamt erzeugte Strukturen: 12
 Davon ausgewählte Strukturen: 1
 $\text{set}\left(\left(\left(\text{np-comp } \textit{der.2}\right) \left(\text{np-comp } \textit{die.3 jungen.1 Schüler.1}\right)\right)\right)$

Eingabe: *der die jungen Schüler unterrichtende*
 Insgesamt erzeugte Strukturen: 26
 Davon ausgewählte Strukturen: 1
 $\text{set}\left(\left(\left(\left(\text{np-open } \textit{der.3}\right) \left(\text{np-comp } \textit{die.3 jungen.1 Schüler.1}\right) \textit{unterrichtende.1}\right)\right)\right)$

Eingabe: *der die jungen Schüler unterrichtende Lehrer*
 Insgesamt erzeugte Strukturen: 50
 Davon ausgewählte Strukturen: 1
 $\text{set}\left(\left(\left(\left(\text{np-comp } \textit{der.3}\right) \left(\text{np-comp } \textit{die.3 jungen.1 Schüler.1}\right) \textit{unterrichtende.1 Lehrer.1}\right)\right)\right)$

Eingabe: *der die jungen Schüler unterrichtende Lehrer schreibt*
 Insgesamt erzeugte Strukturen: 162
 Davon ausgewählte Strukturen: 1
 $\text{set}\left(\left(\left(\left(\left(\text{vp-comp}\right) \left(\text{np-comp } \textit{der.3}\right) \left(\text{np-comp } \textit{die.3 jungen.1 Schüler.1}\right) \textit{unterrichtende.1 Lehrer.1}\right) \textit{schreibt.1}\right)\right)\right)$

Abbildung 4: Heuristische Auswahl der Ergebnisstrukturen

5 PP-Anbindung und die Grenzen des symbolischen Ansatzes

5.1 Das Problem der PP-Anbindung

Beim Parsing mit Wortagenten wird die Entscheidung, ob eine Präpositionalphrase an das Verb bzw. an eine vorausgehende Nominalphrase angebunden wird, u. a. nach folgenden Kriterien getroffen (vgl. [HMS 94]):

- Kann mit Hilfe der Subkategorisierungsinformation beim Verb die Präpositionalphrase angebunden werden? (vgl. Beispiele (1) und (3) in Abbildung 5)
- Kann mit Hilfe der Subkategorisierungsinformation beim Nomen⁷ die Präpositionalphrase angebunden werden? (vgl. Beispiele (2) und (3))
- Kann die Präpositionalphrase mit Hilfe der beim Verb bzw. beim Nomen angegebenen Subkategorisierungsinformation nicht angebunden werden, so wird eine Anbindung an das Verb bzw. an das Nomen mit Hilfe der bei der Präposition verzeichneten Interpretationsinformation versucht. (vgl. Beispiel (4))

Die Anwendung der genannten Entscheidungskriterien führt in zahlreichen Fällen zu eindeutigen Interpretationen. So gelingt die eindeutige Anbindung der durch *über* eingeleiteten Präpositionalphrase in den Beispielen 1 und 2 aufgrund der vorhandenen Subkategorisierung einer *über*-PP bei *schreiben* bzw. *Buch* und der gleichzeitig fehlenden Subkategorisierung bei *Woche* bzw. *verschenken*.

(1) *Der Reporter schrieb in der letzten Woche über das Rockfestival.*

$$\text{set} \left(\left(\left(\left(\text{vp-comp (np-comp der.3 Reporter.1)} \right) \right) \right) \right) \left(\left(\left(\text{schrieb.1} \right) \right) \right) \left(\left(\left(\text{(np-comp in.1 der.3 letzten.1 Woche.1)} \right) \right) \right) \left(\left(\left(\text{(pp-comp über.2 (np-comp das.3 Rockfestival.1))} \right) \right) \right)$$

(2) *Der Lehrer verschenkte das Buch über algebraische Strukturen.*

$$\text{set} \left(\left(\left(\left(\left(\text{vp-comp (np-comp der.3 Lehrer.1)} \right) \right) \right) \right) \right) \left(\left(\left(\text{verschenkte.2} \right) \right) \right) \left(\left(\left(\text{(np-comp das.3 Buch.1 (pp-comp über.2 (np-comp algebraische.1 Strukturen.1))} \right) \right) \right)$$

(3) *Der Autor schrieb den Bericht über alternative Energieformen.*

$$\text{set} \left(\left(\left(\left(\left(\text{vp-comp (np-comp der.3 Autor.1)} \right) \right) \right) \right) \right) \left(\left(\left(\text{schrieb.1} \right) \right) \right) \left(\left(\left(\text{(np-comp den.3 Bericht.1)} \right) \right) \right) \left(\left(\left(\text{(pp-comp über.2 (np-comp alternative.1 Energieformen.1))} \right) \right) \right) \left(\left(\left(\left(\left(\text{vp-comp (np-comp der.3 Autor.1)} \right) \right) \right) \right) \right) \left(\left(\left(\text{schrieb.1} \right) \right) \right) \left(\left(\left(\text{(np-comp den.3 Bericht.1 (pp-comp über.2 (np-comp alternative.1 Energieformen.1))} \right) \right) \right)$$

(4) *Der Künstler malte das Bild über die historischen Ereignisse.*

$$\text{set} \left(\left(\left(\left(\left(\text{vp-comp (np-comp der.3 Künstler.1)} \right) \right) \right) \right) \right) \left(\left(\left(\text{malte.1} \right) \right) \right) \left(\left(\left(\text{(np-comp das.3 Bild.1)} \right) \right) \right) \left(\left(\left(\text{(pp-comp über.2 (np-comp die.3 historischen.1 Ereignisse.1))} \right) \right) \right) \left(\left(\left(\left(\left(\text{vp-comp (np-comp der.3 Künstler.1)} \right) \right) \right) \right) \right) \left(\left(\left(\text{malte.1} \right) \right) \right) \left(\left(\left(\text{(np-comp das.3 Bild.1 (pp-comp über.2 (np-comp die.3 historischen.1 Ereignisse.1))} \right) \right) \right)$$

Abbildung 5: Anbindung von Präpositionalphrasen

In anderen Fällen bleibt trotz vorhandener Subkategorisierungsinformation die Anbindung mehrdeutig. So kann in Beispiel (3) die Präpositionalphrase *über alternative Energieformen* aufgrund

⁷Dabei handelt es sich jeweils um das Head-Nomen einer vorausgehenden Nominalphrase.

der Subkategorisierungsinformation von *schreiben* und *Bericht* sowohl dem Verb als auch dem Nomen untergeordnet werden. Die Auflösung dieser Mehrdeutigkeit ist mit Hilfe des verwendeten symbolischen Ansatzes nicht möglich.⁸

In Beispiel (4) gelingt die Anbindung an das Verb *malen* sowie an das Nomen *Bild* aufgrund der bei *über* eingetragenen Interpretationsinformation (vgl. Abbildung 6).⁹ Wie für Beispiel (3) gibt es auch für Beispiel (4) zwei konkurrierende Interpretationen, die von dem symbolischen Auswahlverfahren nicht weiter disambiguiert werden können. Der Unterschied zu Beispiel (3) besteht darin, daß hier die Subkategorisierung von fakultativen Komplementen des Verbs und des Nomens nicht interferieren. In solchen Fällen könnte eine statistisch ermittelte Anbindungspräferenz weiterhelfen.

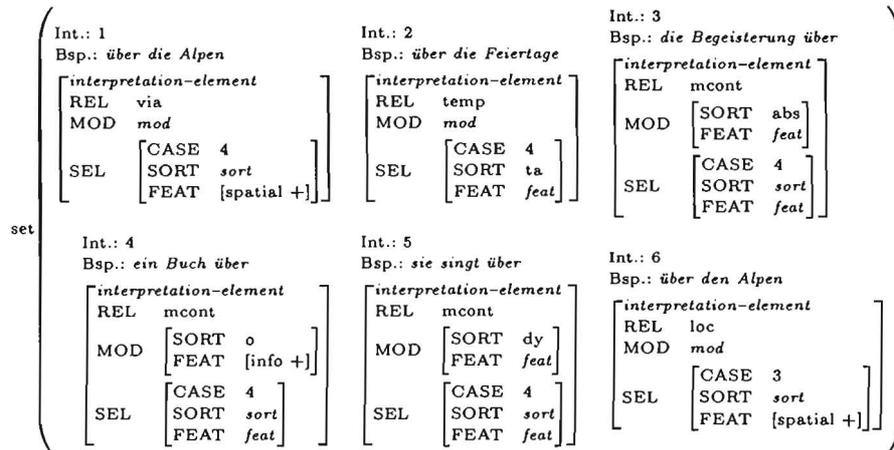


Abbildung 6: Interpretationsinformation zur Präposition *über*

5.2 Mögliche Verbesserungen durch statistische Verfahren

Im Gegensatz zu anderen hybriden Ansätzen (vgl. z. B. [HR 93] für das Englische oder [LMV 97] für das Deutsche), die mit Hilfe von statistischen Verfahren die Anbindungswahrscheinlichkeiten von Präpositionalphrasen u. a. an Nomen und Verben berechnen und diese Ergebnisse in symbolische Analyseverfahren integrieren, wird hier ein anderer Ansatz vorgeschlagen. Der Unterschied besteht darin, daß nicht Anbindungspräferenzen für einzelne Wörter, sondern Präferenzen für bestimmte symbolische Merkmalstrukturen berechnet werden sollen. Diese Strukturen steuern die Anbindung von Präpositionalphrasen mit Hilfe von syntaktischen und semantischen Merkmalen der übergeordneten sowie der untergeordneten Konstituente. Es ist geplant, die relative

⁸Inwieweit es semantisch gesehen ein Unterschied ist, ob die *über*-PP (in der Bedeutungsrelation mcont für "geistiger Inhalt") an das Verb oder an das Nomen angebinden wird, soll hier nicht weiter diskutiert werden. Es ist auch fraglich, ob hier mit Hilfe von statistischen Verfahren eine Präferenz nachzuweisen ist. Beim Wortagentenparsing ist beabsichtigt, solche Mehrdeutigkeiten mit Hilfe des Rechtsassoziationsprinzip zu disambiguieren. Dies kann mit Hilfe des heuristischen Verfahrens insofern simuliert werden, daß in den genannten Fällen Verbalphrasen, die eine geringere Zahl an komplettierten Phrasen enthalten, bevorzugt werden.

⁹Die einzelnen Interpretationen enthalten unter dem Merkmal REL die semantische Relation, in der die beteiligten Konstituenten stehen, z. B. lokale Erstreckung (via), Zeitangabe (temp), geistiger Inhalt (mcont) und Lokation (loc). Das Merkmal MOD beschränkt die Merkmalwerte der übergeordneten Konstituente, das Merkmal SEL die der subkategorisierten Konstituente. Die Merkmale CASE, SORT und FEAT dienen der Angabe von Kasus, von semantischen Sorten bzw. von semantischen Feature-Werten. Wichtige Sorten der zugrundeliegenden Ontologie sind beispielsweise konkrete (co) und abstrakte (ab) Objekte (o), abstrakte Sachverhalte (abs), statische Situationen bzw. Zustände (st) und dynamische Situationen bzw. Vorgänge (dy). Letztere gliedern sich noch weiter in Handlungen (da) und Geschehnisse (dn). Feature-Werte unter dem Merkmal FEAT dienen der zusätzlichen semantischen Einschränkung, z. B. [legper +] für "legale Person", [spatial -] für "hat räumliche Ausdehnung" oder [info +] für Informationsträger (vgl. dazu [HS 97]). Die in Abbildung 6 angegebenen Interpretationsnummern und Beispiele dienen lediglich der Veranschaulichung.

Häufigkeit der Anwendung einer solchen Struktur mit Hilfe von Korpusuntersuchungen festzustellen.

Wie ein solches hybrides Verfahren aussehen könnte, soll am Beispiel der Präposition *über* verdeutlicht werden. Die in Abbildung 6 angegebene Interpretationsinformation zu *über* zeigt, daß in zahlreichen Fällen das MOD-Merkmal unterspezifiziert ist (vgl. die Angabe des Typs *mod*). In diesen Fällen ist die Anbindung der betreffenden Präpositionalphrase an das Verb sowie an ein vorausgehendes Nomen möglich. In den anderen Fällen ist durch das MOD-Merkmal die Anbindung an das Nomen (vgl. z. B. Int. 3 und 4) bzw. an das Verb (vgl. Int. 5) vorgegeben.¹⁰ Um nun eine Aussage über die Anbindungspräferenz der Präposition machen zu können, werden die unterspezifizierten Sortenangaben unter MOD in die folgenden disjunkten Sortenangaben aufgeteilt:¹¹ bei den Nomen *co* für konkrete und *ab* für abstrakte Objekte und bei den Verben *st* für statische und *dy* für dynamische Situationen.

Int.: 1 Bsp.: <i>die Brücke über den Fluß</i>	Int.: 2 Bsp.: <i>der Flug über die Alpen</i>	Int.: 3 Bsp.:	Int.: 3 Bsp.: <i>er fliegt über die Alpen</i>
$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL via} \\ \text{MOD [SORT co]} \\ \text{SEL [CASE 4} \\ \text{SORT sort} \\ \text{FEAT [spatial +]} \\ \text{VAL p(int}_1\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL via} \\ \text{MOD [SORT ab]} \\ \text{SEL [CASE 4} \\ \text{SORT sort} \\ \text{FEAT [spatial +]} \\ \text{VAL p(int}_2\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL via} \\ \text{MOD [SORT st]} \\ \text{SEL [CASE 4} \\ \text{SORT sort} \\ \text{FEAT [spatial +]} \\ \text{VAL p(int}_3\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL via} \\ \text{MOD [SORT dy]} \\ \text{SEL [CASE 4} \\ \text{SORT sort} \\ \text{FEAT [spatial +]} \\ \text{VAL p(int}_4\text{)} \end{array} \right]$
Int.: 5 Bsp.:	Int.: 6 Bsp.: <i>der Urlaub über Ostern</i>	Int.: 7 Bsp.: <i>er erkrankt über Ostern</i>	Int.: 8 Bsp.: <i>sie fährt über Pfingsten</i>
$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL temp} \\ \text{MOD [SORT co]} \\ \text{SEL [CASE 4} \\ \text{SORT ta} \\ \text{FEAT feat} \\ \text{VAL p(int}_5\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL temp} \\ \text{MOD [SORT ab]} \\ \text{SEL [CASE 4} \\ \text{SORT ta} \\ \text{FEAT feat} \\ \text{VAL p(int}_6\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL temp} \\ \text{MOD [SORT st]} \\ \text{SEL [CASE 4} \\ \text{SORT ta} \\ \text{FEAT feat} \\ \text{VAL p(int}_7\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL temp} \\ \text{MOD [SORT st]} \\ \text{SEL [CASE 4} \\ \text{SORT ta} \\ \text{FEAT feat} \\ \text{VAL p(int}_8\text{)} \end{array} \right]$
Int.: 9 Bsp.: <i>die Begeisterung über</i>	Int.: 10 Bsp.: <i>ein Buch über</i>	Int.: 11 Bsp.: <i>sie singt über</i>	Int.: 12 Bsp.: <i>das Flugzeug über den Alpen</i>
$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL mcont} \\ \text{MOD [SORT abs} \\ \text{FEAT feat} \\ \text{SEL [CASE 4} \\ \text{SORT sort} \\ \text{FEAT feat} \\ \text{VAL p(int}_9\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL mcont} \\ \text{MOD [SORT o} \\ \text{FEAT [info +]} \\ \text{SEL [CASE 4} \\ \text{SORT sort} \\ \text{FEAT feat} \\ \text{VAL p(int}_{10}\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL mcont} \\ \text{MOD [SORT dy]} \\ \text{SEL [CASE 4} \\ \text{SORT sort} \\ \text{FEAT feat} \\ \text{VAL p(int}_{11}\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL loc} \\ \text{MOD [SORT co]} \\ \text{SEL [CASE 3} \\ \text{SORT sort} \\ \text{FEAT [spatial +]} \\ \text{VAL p(int}_{12}\text{)} \end{array} \right]$
Int.: 13 Bsp.: <i>der Flug über den Bergen</i>	Int.: 14 Bsp.: <i>er erkrankt über den Alpen</i>	Int.: 15 Bsp.: <i>es fliegt über dem Haus</i>	Int.: 16 Bsp.:
$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL loc} \\ \text{MOD [SORT ab]} \\ \text{SEL [CASE 3} \\ \text{SORT sort} \\ \text{FEAT [spatial +]} \\ \text{VAL p(int}_{13}\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL loc} \\ \text{MOD [SORT st]} \\ \text{SEL [CASE 3} \\ \text{SORT sort} \\ \text{FEAT [spatial +]} \\ \text{VAL p(int}_{14}\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL loc} \\ \text{MOD [SORT dy]} \\ \text{SEL [CASE 3} \\ \text{SORT sort} \\ \text{FEAT [spatial +]} \\ \text{VAL p(int}_{15}\text{)} \end{array} \right]$	$\left[\begin{array}{l} \text{interpretation-element} \\ \text{REL rel} \\ \text{MOD mod} \\ \text{SEL sel} \\ \text{VAL p(int}_{16}\text{)} \end{array} \right]$

Abbildung 7: Expandierte Interpretationsinformation zu *über*

Diese Aufteilung hat eine Erhöhung der Merkmalstrukturen in der Interpretationsinformation von *über* zur Folge. In Abbildung 7 ist die neue Menge von Merkmalstrukturen angegeben. Zusätzlich ist dort eine "Dummy"-Struktur (vgl. Int. 16) eingetragen, die immer dann zur Anwendung kommt, wenn bei der Disambiguierung keine der anderen Merkmalstrukturen angewendet wird. Die relative Häufigkeit dieser Struktur kann dann als ein Maß für die Abdeckung

¹⁰Die semantische Sorte *dy* bezieht sich ausschließlich auf Verben, die Sorten *abs* und *o* ausschließlich auf Nomen.

¹¹Damit bleiben die Sortenangaben allerdings immer noch unterspezifiziert. Es soll aber zunächst anhand dieser relativ allgemeinen Sorten versucht werden, statistische Anbindungspräferenzen zu ermitteln.

der vorhandenen Interpretationsinformation herangezogen werden. Bei der geplanten Korpusuntersuchung sollen dann die relativen Häufigkeiten der Stukturanwendungen bei den betreffenden Merkmalstrukturen verzeichnet werden (vgl. Merkmal VAL).

Ein wichtiger Vorteil des hier vorgeschlagenen Ansatzes ist, daß er semantische Information in die statistischen Untersuchungen einbezieht.¹² Eine Orientierung z. B. der np_attach-Werte an den Wortformen der Nomen, wie es u. a. von [LMV 97] vorgeschlagen wird, kann große Probleme mit sich bringen. Man denke nur daran, daß u. U. für jede einzelne Wortform eines Lexems ein Eintrag in der Präferenzliste gemacht werden muß. Darüber hinaus werden z. B. auch Eigennamen in die Liste eingetragen, falls es im Korpus dafür statistische Evidenz gibt (z. B. *Peter aus Hamburg, Paul aus München*). Das Beispiel der Eigennamen verdeutlicht auch, daß die Generalisierung auf syntaktische und semantische Merkmale durchaus sinnvoll ist.

Literatur

- [Eim 88] Michael Eimermacher: Wortorientiertes Parsen. Dissertation. Fachbereich Informatik. Technische Universität Berlin, 1988
- [Han 91] Gerhard Hanrieder: Robustes Wortparsing. Lexikonbasierte morphologische Analyse (komplexer) deutscher Wortformen. Magisterarbeit, Universität Trier, Februar 1991
- [Han 95] Gerhard Hanrieder: Kognitiv adäquates Parsing mit LAUG. In: Ludwig Hitzenberger (Hrsg.): *Angewandte Computerlinguistik*. Hildesheim: Georg Olms Verlag 1995, S. 93–110
- [Hau 89] Roland Hausser: *Computation of Language. An Essay on Syntax, Semantics and Pragmatics in Natural Man-Machine Communication*. Berlin, New York: Springer 1989
- [Hem 93] Barbara Hemforth: *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens. Dissertationen zur Künstlichen Intelligenz, Bd. 40*. Sankt Augustin: infix 1993
- [HM 94] Hermann Helbig, Andreas Mertens: Word Agent Based Natural Language Processing. In: Loe Boves, Anton Nijholt (eds.): *Speech and Language Engineering, Proceedings of the 8th Twente Workshop on Language Technology*, Enschede, NL: Faculteit Informatica, Universiteit Twente 1994, S. 65–74
- [HMS 94] Hermann Helbig, Andreas Mertens, Marion Schulz: Die Rolle des Lexikons bei der Disambiguierung. In: Harald Trost (Hrsg.): *KONVENS '94 – Verarbeitung natürlicher Sprache*, Wien, 28.–30. September 1994. Berlin 1994, S. 151–160
- [HR 93] Donald Hindle, Mats Rooth: Structural Ambiguity and Lexical Relations. In: *Computational Linguistics*, 19 (1), 1993, S. 103–120
- [HS 97] Hermann Helbig, Marion Schulz: Knowledge Representation with MESNET – A Multilayered Extended Semantic Network. Working Notes of the 1997 AAAI Spring Symposium on Ontological Engineering, Stanford CA 1997, S. 64–72
- [HSB 94] Udo Hahn, Susanne Schacht, Norbert Bröker: Concurrent, Object-Oriented Dependency Parsing: The *ParseTalk* Model. In: *International Journal of Human-Computer Studies*, 41, 1994 (1/2), S. 179–222
- [LMV 97] Hagen Langer, Stephan Mehl, Martin Volk: Hybride NLP-Systeme und das Problem der PP-Anbindung. In: *Elektronische Proceedings des Workshops 03 "Hybride konnektionistische, statistische und symbolische Verarbeitung natürlicher Sprache"* auf der 21. Deutschen Jahrestagung für Künstliche Intelligenz 1997, Freiburg (forthcoming)
- [MC 94] David Milward, Robin Cooper: Incremental Interpretation: Applications, Theory, and Relationship to Dynamic Semantics. In: *Proceedings of the 15th International Conference on Computational Linguistics, COLING 94*, Kyoto, Japan, S. 748–754
- [Mer 97] Andreas Mertens: Robustes Parsing mit Wortagenten. In: *Proceedings der 10. GLDV Jahrestagung*, Leipzig 17.–19.03.1997 (forthcoming)
- [Sma 87] Steven Small: A Distributed Word-Based Approach to Parsing. In: L. Bolc (ed.): *Natural Language Parsing Systems*. Berlin: Springer 1987, S. 161–201

¹²Dies ist allerdings auch ein Problem, denn für repräsentative Untersuchungen muß das verwendete Lexikon die semantischen Sorten und Feature-Werte der vorkommenden Verben und Nomen enthalten.

Hybride NLP-Systeme und das Problem der PP-Anbindung

Hagen Langer (Universität Osnabrück)

hlanger@cl-ki.uni-osnabrueck.de

Stephan Mehl (Universität Duisburg)

he234me@unidui.uni-duisburg.de

Martin Volk (Universität Zürich)

volk@ifl.unizh.ch

1 Problemstellung

Das Problem der Anbindungsambiguitäten bei Präpositionalphrasen ist zwar bereits oft und unter verschiedenen Aspekten - von der Sprachtechnologie bis hin zur Psycholinguistik - untersucht worden, es darf aber gleichwohl als nach wie vor ungelöst angesehen werden. PP-Anbindung ist ein zentrales Problem, da es sich bei Präpositionalphrasen um alles andere als ein marginales linguistisches Phänomen handelt: In deutschen Zeitungstexten kommt auf einen Satz im Durchschnitt etwa eine Präpositionalphrase, in fachsprachlichen Texten kann dieser Wert noch erheblich höher liegen¹.

Die bei Präpositionalphrasen auftretenden Disambiguierungsprobleme sind vielfältig und komplex (vgl. [MME96] für einen Überblick). In Systemen, die ausschließlich auf symbolischen Repräsentationen grammatischen Wissens beruhen, wird das Problem der PP-Anbindung nicht selten "ausgeklammert" (vgl. [Wau96], [RS96]), d.h. es wird auf den Versuch der Disambiguierung systematisch verzichtet, um der daraus resultierenden kombinatorischen Explosion von Lesarten zu entgehen. Diese Strategie wäre sinnvoll, wenn es semantische Interpretationsverfahren gäbe, die die erforderliche Disambiguierung mit einem geringen Aufwand leisten könnten und eine syntaktische Vorverarbeitung überhaupt nicht oder nur partiell voraussetzten. Solche Verfahren sind jedoch bestenfalls für thematisch äußerst eingeschränkte Textsorten vorstellbar. In der letzten Zeit sind daher verschiedene statistische Verfahren als Ergänzung zu klassischen Parser-Architekturen vorgeschlagen worden ([HR93], [EC96], [Fra96], [BR94], [HangLi96]). Allen diesen Verfahren ist gemeinsam, dass sie anhand möglichst großer Korpora ([EC96] verwendet ein Korpus mit 36 Mio. Wortformen) ermitteln, wie hoch die Wahrscheinlichkeit ist, dass eine PP an ein bestimmtes Nomen, Verb oder Adjektiv angebunden wird. Einerseits handelt es sich also um Verfahren zur statistischen Bestimmung von Valenzrahmen (in einem weiten Sinn; auch nicht-deverbative Nomina können eine so enge statistisch belegbare Bindung an eine PP aufweisen, dass solche Attribute mit Valenzstellen verglichen werden können, wie etwa *Familien mit Kindern*). Andererseits werden diese quantitativ bewerteten Valenzstellen als Parsingpräferenzen interpretiert. Das heißt, dass bei Zuordnungsmehrdeutigkeiten dasjenige Lexem als Matrixkopf ausgewählt wird, das die stärkere Valenzstelle aufweist. So ist z.B. in den beiden folgenden Beispielsätzen die 'mit'-PP hinsichtlich ihrer Anbindung an Nomen oder Verb mehrdeutig:

*Seit 1982 arbeitete das Standardisierungs-Gremium
in enger Zusammenarbeit mit IEEE an den Basisnormen.*

¹Eine Stichprobe von 65 sehr kurzen Sätzen (durchschnittliche Satzlänge: 11.69 Wörter) aus einem Korpus von wissenschaftlichen Abstracts ergab z.B. bereits einen Wert von 1.2 PPs pro Satz; lediglich 11 der 65 Sätze enthielten überhaupt keine PP.

Zusammenarbeit - mit: Präferenz² = 0.4452174
arbeiten - mit: Präferenz = 0.38

*Die Computer arbeiten auch als Server mit Microsofts Hochleistungsbetriebssystem
Windows NT.*

Server - mit: Präferenz = 0.0386905
arbeiten - mit: Präferenz = 0.38

Von den erwähnten Arbeiten erzielten [EC96] auf der Grundlage eines 36-Millionen-Korpus die besten Resultate, und zwar sowohl bei der Valenzbestimmung als auch bei der Disambiguierung. So weisen die von ihnen ermittelten Valenzangaben bei den Verben eine Präzision von 92% und einen Recall von 52% auf. Die mit diesen Valenzstellen verbundenen Bewertungen steigerten die Analysegenauigkeit des von ihnen verwendeten probabilistischen Parsers von 84.9% auf 91.1%, d.h., die Fehlerrate sank um 41.1%!

Nichtsdestotrotz wirft auch diese Arbeit ebenso wie die übrigen noch zahlreiche Fragen auf:

1. Sind die bislang nur für das Englische geprüften Verfahren auch auf eine Sprache wie das Deutsche übertragbar, das bekanntlich trotz genealogischer Verwandtschaft mit dem Englischen gerade unter Aspekten des Parsings völlig andere Probleme aufwirft (sog. "freie" Wortstellung, long-distance-dependencies usw.)? Im Hinblick auf das PP-Zuordnungsproblem spielt dabei eine besondere Rolle, dass NPs im Genitiv oder Dativ im Englischen durch eine PP wiedergegeben werden. Den sehr häufigen Genitivattributen entsprechen also im Englischen PPs mit Nomenzuordnung. Während Verb- und Nomenzuordnungen im Deutschen tendenziell etwa gleich häufig vorkommen (mit textsortenspezifischen Schwankungen), dominieren im Englischen daher die Nomenzuordnungen (65% bei [EC96]). [HR93] nutzen dies, um die Nomenzuordnung in ihrer Zuordnungsheuristik als Default zu verwenden. – Ein weiteres schwerwiegendes Problem besteht darin, dass Verfahren wie das von Hindle und Rooth vorgeschlagene prinzipiell ein vollständiges Substantiv-Lexikon voraussetzen. Eine solche Voraussetzung ist für Sprachen, in denen nominale Ad-hoc-Komposition - wie im Deutschen - hochproduktiv ist, natürlich nicht erfüllt.
2. Die schwächste Stelle der bisherigen Arbeiten liegt in der Unvollständigkeit der ermittelten Valenzangaben. Beruhen die noch verbleibenden Analysefehler auf dieser Unvollständigkeit, oder handelt es sich um Fälle, die grundsätzlich nur mit einer semantischen Feinanalyse aufgelöst werden können?
3. Welche Rolle spielen fachsprachliche Unterschiede? Die meisten bisherigen Studien beruhen überwiegend auf Wall-Street-Journal-Texten, für die in der Penn Treebank fertige Analysen zur Evaluierung vorliegen. Sind die ermittelten Ergebnisse auch auf allgemeinsprachliche Korpora oder andere Fachsprachen übertragbar?

In Abschnitt 2 geben wir zunächst einen knappen Überblick über die Möglichkeiten und Grenzen traditioneller Verfahren. In Abschnitt 3 berichten wir über einige experimentelle Resultate, die wir mit kombinierten symbolischen und statistischen Techniken erzielt haben, und versuchen schließlich in Abschnitt 4 daraus Konsequenzen abzuleiten, die zu einer Beantwortung der genannten Fragen beitragen können.

2 Bisherige Ansätze

Die systematische Ausnutzung von lexikalischen Kookkurrenzen zwischen Präpositionen, einerseits, und Verben bzw. Nomina, andererseits, wurde erstmals von [HR93] systematisch verfolgt und u.a.

²Diese Werte wurden aus einem Korpus von Artikeln aus der Computer Zeitung (Konradin Verlag) mit einem Umfang von 2,5 Millionen Wörtern erhoben. Der Präferenzwert eines Wortes ist hier als der Anteil seiner Vorkommen definiert, bei denen es mit einer 'mit'-PP vorkommt.

von den oben genannten Autoren weiterentwickelt.

Eine möglichst exakte Ermittlung von Konkurrenzpräferenzen würde die folgenden Schritte erfordern:

- lexikalische Disambiguierung von Verben, Nomina und Präpositionen
- syntaktische Disambiguierung der PP-Zuordnung
- Berechnung der Konkurrenzwahrscheinlichkeiten mindestens für die Tripel <Verb, Nomen, Präposition>, wenn nicht für die Quadrupel <Verb, Nomen, Präposition, regiertes Nomen> oder sogar für die n-Tupel, die sich daraus ergeben, dass mehrere Nomina als Anknüpfungspunkte zur Verfügung stehen.

Selbstverständlich ist der letztere, aber auch die beiden ersten Punkte undurchführbar. Für das Deutsche werden mehr als 100 verschiedene Präpositionen angenommen, für Verben und Nomina liegen solche Schätzungen nicht vor - beide Klassen werden als offen angenommen. Wenn man also lediglich die 1000 häufigsten Nomina und die 1000 häufigsten Verben in eine Analyse einbeziehen möchte und jeweils mindestens 30 Belege pro Tripel ansetzt, ergibt sich eine theoretische Gesamtzahl von 3.000.000.000 Sätzen, die manuell beurteilt werden müssten.

Stattdessen betrachten alle Autoren die konkurrierenden Präferenzen für die PP-Zuordnung als voneinander unabhängig, und alle Lesarten aller beteiligten Lexeme werden zusammengeworfen. Strukturell disambiguierte Sätze (d.h. eine Treebank) werden nur zur Evaluierung benutzt, weil weder für das Englische noch für das Deutsche hinreichend große Treebanks zur Verfügung stehen. Abstrahiert man von verschiedenen Verfeinerungen der verwendeten Methoden, so wird gemessen, wie oft ein bestimmtes Lexem mit einer bestimmten PP zusammen vorkommt, in Relation zu seiner Okkurrenz ohne eine solche PP. Diese Zahlen sind also durch die folgenden Faktoren verrauscht:

- durch Homographie und Polysemie
- durch Fälle, in denen eine PP tatsächlich dem betreffenden Lexem zugeordnet ist, aber in anderer Funktion, und
- vor allem durch Fälle, in denen eine ganz andere Zuordnung vorliegt.

Der letzte Fehler läßt sich nur dann gering halten, wenn hinreichend große Korpora verwendet werden und wenn hinreichend viele Belege syntaktisch eindeutig sind (z.B. bei kurzen Sätzen, wie sie die Mehrheit in allgemeinsprachlichen Korpora bilden). Umgekehrt ist in allgemeinsprachlichen Korpora das Problem der Polysemie wesentlich höher als in fachsprachlichen.

Aufgrund dieser zahlreichen Fehlerquellen funktioniert das Verfahren nur für häufig belegte Lexeme. Aus diesem Grund verwenden [EC96] bei der Evaluation ihrer Ergebnisse nur solche Belege, bei denen die beteiligten Lexeme mindestens 500mal im Korpus vertreten sind.

3 Eigene Untersuchungen

Da für das Deutsche weder derartig umfangreiche Korpora noch Treebanks zur Verfügung stehen, mußten wir uns bisher auf eher stichprobenartige Analysen mit einem hohen Anteil an manueller Evaluierung beschränken, um uns einer Antwort auf die oben gestellten Fragen zu nähern. Dies begann mit der Auswahl einer exemplarischen Präposition, und zwar 'mit'. Neben dem pragmatischen Grund, dass 'mit' zu den häufigsten Präpositionen des Deutschen gehört und somit auch bei kleineren Korpora hinreichend viele Daten zu erwarten sind, bietet sich 'mit' als Testinstanz auch deshalb an, weil es massiv polysem ist (instrumental, komitativ, partitiv usw.) und in zahlreichen syntaktischen Funktionen vorkommen kann:

- Präpositionalobjekt: er rechnet mit seiner Ankunft
- Attribut zu einem Nomen: der Mann mit dem Hut

Umgang	57.171930
Interview	33.244604
Zusammenarbeit	27.737401
Zusammenhang	26.025830
Begegnung	24.948718
Begegnungen	23.843137
Beschaeftigung	23.161905
...	...
Jahre	0.313770
Ende	0.260450

Tabelle 1: Präferenzen bei Nomina

- Attribut zu einem Adjektiv: er war mit dem Resultat zufrieden
- Komplement eines Nomens: der Umgang mit den Leuten
- Adverbial: er kam mit 5 Minuten Verspätung
- Bestandteil von idiomatischen Wendungen: mit Fug und Recht usw.

3.1 Experiment 1: Kookkurrenzpräferenzen bei Nomina

Aus einem Mischkorpus³ mit 1,1 Millionen Wörtern extrahierten wir alle Sätze, in denen 'mit' als orthographisches Wort vorkam. Die Extraktion erfolgte vollautomatisch und ist relativ verlässlich⁴. Diese Teilmenge enthielt 8599 Sätze. Aus dieser Satzmenge ermittelten wir mit einem Pattern-Matcher für reguläre Ausdrücke alle Nomina N_i (i.e. großgeschriebene Wörter), die links vor einem 'mit' vorkamen und zählten diese Vorkommen $Anz(N_i)$, wodurch sich eine relative Häufigkeit $p_m(N_i) = \frac{Anz(N_i)}{\sum_{j=1, n} Anz(N_j)}$ angeben ließ. Außerdem ermittelten wir die relative Häufigkeit $p(N_i)$ aller Nomina im Gesamtkorpus und definierten mit diesen beiden Werten das Maß für NP-Anbindungspräferenz⁵ eines Nomens $np_attach(N_i) = \frac{p_m(N_i)}{p(N_i)}$. Dieses Datenakquisitionsverfahren hat folgende systematische Schwächen:

- das Bezugsnomen einer PP ist im Deutschen nicht immer das *unmittelbare* Vorgängernomen: *die Verbindung der Computer mit einem Koaxialkabel*
- das Bezugsnomen kann auch rechts der PP stehen: *er steht mit ihm in Verbindung*

Dennoch ergaben sich mit diesem Verfahren Werte, die sich mit unseren Intuitionen deckten; Beispiele für die np_attach -Werte einiger Nomina sind in Tabelle 1 aufgeführt.

Wir wählten aus dieser Liste mit insgesamt 65.500 Einträgen eine Teilmenge von 73 Nomina mit signifikant hohen bzw. signifikant niedrigen np_attach -Werten aus und integrierten diese Zusatzinformation in den GEPARD-Parser. Das GEPARD-System ist ein in C implementierter

³Das Korpus enthält zu etwa 80% Zeitungsartikel, aber auch wissenschaftliche Abstracts, Ausschnitte aus literarischen Texten und einen kleinen Anteil spontansprachlicher Äußerungen.

⁴Die Extraktion scheitert in Einzelfällen, z.B. kann das Vorkommen von Ziffern vor einem Punkt am Satzende nicht von einer Ordinalzahl unterschieden werden. Solche Ungenauigkeiten sind für unsere Fragestellung aber nicht von essentieller Bedeutung, weswegen wir für dieses Experiment auf eine nachträgliche manuelle Korrektur verzichteten.

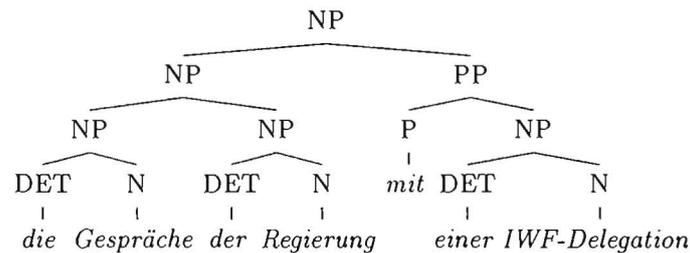
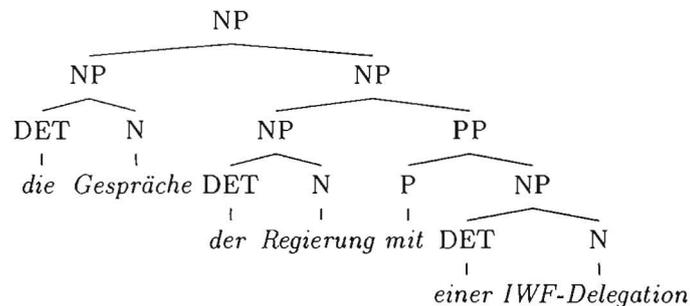
⁵Wenn der Wert für ein Nomen bei 1 liegt, heißt dies, dass es in dem Kontext von mit-PPs weder seltener noch häufiger vorkommt als in anderen Kontexten. Höhere Werte sind ein Indiz dafür, dass für das Nomen eine überdurchschnittliche Tendenz feststellbar ist, im Kontext von mit-PPs vorzukommen. Dieses Maß hat den neutralen Wert 1 und kann somit als Multiplikator mit anderen Werten, z.B. Wahrscheinlichkeiten von Strukturbäumen, verrechnet werden.

Chart-Parser, der von einem mehrstufigen Parser-Compiler aus einer kontextfreien Grammatik mit Merkmalsannotationen generiert wurde⁶. Das System umfasst gegenwärtig etwa 2000 Syntaxregeln und ein Vollformenlexikon mit 125.000 Einträgen. Die kontextfreien Regeln sind mit Wahrscheinlichkeiten annotiert, die aus einem Korpus von 13000 Testsätzen ermittelt wurden. Trotz einer relativ detaillierten Behandlung von Präpositionalkonstruktionen in der zugrundeliegenden Grammatik (vgl. [Lan96]) kann der GEPARD-Parser bestimmte Anbindungsambiguitäten prinzipiell nicht auflösen, z.B. solche Fälle, bei denen innerhalb einer komplexen NP mehrere Anbindungsmöglichkeiten für eine attributive PP vorliegen. Genau für diesen Ambiguitätstyp liefern die np_attach-Werte ein Entscheidungskriterium. Die Berechnung der Gesamtwahrscheinlichkeit einer Strukturbeschreibung wurde wie folgt modifiziert: Bei jeder Ableitung, in der eine 'mit'-PP an eine NP angebunden wurde, für deren Kopf-Nomen ein np_attach-Wert vorlag, wurde die aus der Bewertung der kontextfreien Grammatik gewonnene Wahrscheinlichkeit mit dem np_attach-Wert des jeweiligen Nomens multipliziert. Sofern der np_attach-Wert für das jeweilige Nomen über 1 liegt, ergibt sich dadurch eine entsprechende Erhöhung der Gesamtwahrscheinlichkeit der Ableitung, bei niedrigeren Werten eine Absenkung.

Anschließend wurde mit dem um die zusätzliche NP-Anbindungsheuristik erweiterten Parser eine Testsatzmenge von 100 Sätzen analysiert. Alle Sätze sind belegte Fälle (zumeist aus der 'Süddeutschen Zeitung') und enthalten jeweils (mindestens) eine 'mit'-PP, bei der die Anbindung an ein Nomen syntaktisch möglich, aber nicht eindeutig ist (andere Ambiguitätstypen wie die Unterscheidung zwischen Präpositionalobjekt und Adverbial wurden in diesem Experiment nicht überprüft). So gibt es z.B. in dem Testsatz:

Angeblich standen die Gespräche der Regierung mit einer IWF-Delegation kurz vor dem Abschluß.

eine adverbiale und zwei NP-Anbindungsmöglichkeiten für die PP *mit einer IWF-Delegation*:



⁶Über <ftp://cl-ki.uni-osnabrueck.de/pub/export/persons/langer> ist eine Demonstratorversion des Parsers via anonymous-ftp zugänglich.

3.2 Experiment 2: Kookkurrenzpräferenzen bei Verben

Den Bereich der Verben haben wir auf der Grundlage von zwei verschiedenen Korpora untersucht: einem fachsprachlichen mit ca. 2.5 Millionen und einem allgemeinsprachlichen (LIMAS-Korpus) mit 1.1 Millionen Wortformen. Die Analyse des fachsprachlichen Korpus ergab 286 Verben, die in mindestens 10 Belegen mit einer 'mit'-PP zusammen auftreten; in dem knapp die Hälfte kleineren LIMAS-Korpus waren es nur noch 51. In den meisten Fällen stellen diese Belege allerdings nur einen Bruchteil der tatsächlichen Vorkommen des betreffenden Verbs dar. Hierfür gibt es zwei Erklärungen:

- Es handelt sich tatsächlich nicht um Valenzstellen, sondern um Adjunkte oder um PPs, die einer NP zugeordnet sind.
- Die Valenzstelle existiert nur bei einer von mehreren Lesarten, die dazu möglicherweise nur selten benutzt wird.

Die beiden Fälle sind statistisch nicht zu unterscheiden. Obwohl mit dieser Methode durchaus einige Verben mit 'mit'-Valenz zu finden sind (z.B. *arbeiten*, *rechnen*, *verhandeln*, *kämpfen*, *verbinden*, *sich befassen*, *sich beschäftigen*), ist eine automatische Akquisition der Valenzangaben ohne erheblich größere Korpora nicht durchführbar.

Besser sieht es dagegen aus, wenn man z.B. aufgrund eines Wörterbuchs vorgibt, welche Verben über fakultative Valenzstellen verfügen, und für diese fakultativen Valenzstellen Präferenzstärken ermittelt. Eine erste Analyse von 182 mehrdeutigen Sätzen für insgesamt 7 Verben ergab, dass nur in 8 Fällen die betreffende PP nicht dem Verb zuzuordnen war; bei 14 weiteren Belegen handelte es sich um ein Adjunkt statt des erwarteten Komplements. Allerdings liegen noch keine Angaben darüber vor, wie häufig die Verbvalenzstellen in diesen Belegen tatsächlich in Konkurrenz zu Valenzstellen treten, die von einem Nomen eröffnet werden. Es ist denkbar, dass diese Disambiguierungserfolge ebenso mit einer nicht quantitativ bewerteten Valenzangabe erzielt werden können. Dies gilt im übrigen auch für die von [EC96] angegebenen Ergebnisse; gerade im Hinblick auf die Evaluierung der Disambiguierung fassen sich [EC96] extrem kurz.

Des Weiteren ist noch zu prüfen, wieweit sich die Unterschiede in der Häufigkeitsverteilung der Verben zwischen dem allgemein- und dem fachsprachlichen Korpus auch in unterschiedlichen Bindungsstärken ausdrücken. Beispielsweise sind *arbeiten mit*, *rechnen mit*, *ausstatten mit* in dem technischen Korpus wesentlich häufiger.

4 Folgerungen und Ausblick

Wir haben die Umsetzung der Grundideen von Hindle & Rooth bzw. von Ersan & Charniak mit einem Modell untersucht, das weder eine Treebank noch eine manuelle Beurteilung der Trainingsdaten voraussetzt und dennoch bei kleinen Stichproben, die wir untersuchen konnten, in relativ vielen Fällen erfolgreich zur Disambiguierung führte. Es zeigte sich allerdings, insbesondere bei den Verben, dass subtilere Unterscheidungen, wie z.B. die Unterscheidung zwischen echten Präpositionalobjekten und solchen Kookkurrenzpräferenzen, die darauf beruhen, dass einige Verben besonders häufig mit bestimmten Adverbialen (z.B. instrumentalen) auftreten, bislang nur sehr bedingt aus Rohkorpora extrahiert werden konnten.

Dies entspricht den Erfahrungen mit statistischen Methoden bei anderen Fragestellungen: So können z.B. HMM-basierte POS-Tagger, die eine Korrektheit von über 95% erreichen, für kleine Tagsets (unter 100 Tags) mit einfachen Mitteln aus kleinen Trainingskorpora gewonnen werden. Für größere Tagsets, die solche letztendlich unentbehrlichen syntaktischen Differenzierungen wie die Unterscheidung zwischen verschiedenen Verbvalenzen beinhalten, sind vergleichbare Ergebnisse jedoch bislang nicht erreichbar.

Die von uns verwendeten statistische Modelle können für differenziertere Disambiguierungsaufgaben noch nicht erfolgreich eingesetzt werden, d.h. man bleibt auch weiterhin auf möglichst detaillierte (symbolische) Repräsentationen der grammatischen und semantischen Eigenschaften

der beteiligten Lexeme und Strukturtypen angewiesen und kann statistische Modelle lediglich als zusätzliche Heuristik verwenden. Ob die Verwendung sehr großer Treebanks oder gezielterer Datenakquisitionsverfahren (wie z.B. die Auswertung der Kovariation der Kookkurrenzpräferenzen von 'mit' und 'ohne' bei Verben) oder eine Kombination davon auf die Dauer zu einem Fortschritt führen, ist somit noch eine offene Frage, deren Beantwortung erst dann in Angriff genommen werden kann, wenn eine Treebank für ein Multi-Millionen-Korpus des Deutschen existiert.

Literatur

- [BR94] E. Brill and P. Resnik. A rule-based approach to prepositional phrase attachment disambiguation. In *Proceedings of COLING*, pages 1198–1204, Kyoto, 1994. ACL.
- [EC96] Murat Ersan and Eugene Charniak. A statistical syntactic disambiguation program and what it learns. In S. Wermter, E. Riloff, and G. Scheler, editors, *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, volume 1040 of *Lecture Notes in AI*, pages 146–159. Springer Verlag, Berlin, 1996.
- [Fra96] Alex Franz. Learning PP attachment from corpus statistics. In S. Wermter, E. Riloff, and G. Scheler, editors, *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, volume 1040 of *Lecture Notes in AI*, pages 188–202. Springer Verlag, Berlin, 1996.
- [HR93] D. Hindle and M. Rooth. Structural ambiguity and lexical relations. *Computational Linguistics*, 19(1):103–120, 1993.
- [Lan96] Hagen Langer. Disambiguierung von Präpositionalkonstruktionen mit einem syntaktischen Parser: Möglichkeiten und Grenzen. In S. Mehl, A. Mertens, and M. Schulz, editors, *Präpositionalsemantik und PP-Anbindung*, number SI-16 in Schriftenreihe Informatik, pages 23–31. Gerhard-Mercator-Universität, Gesamthochschule Duisburg, 1996.
- [HangLi96] Hang Li. A Probabilistic Disambiguation Method Based on Psycholinguistic Principles. In *Proc. 4th Workshop on Very Large Corpora*, pages 141–154. 1996.
- [MME96] S. Mehl, A. Mertens, and M. Schulz (Eds.). *Präpositionalsemantik und PP-Anbindung*. Schriftenreihe Informatik SI-16, Gerhard-Mercator-Universität, GH Duisburg, 1996.
- [RS96] F. Richter and M. Sailer. Syntax für eine unterspezifizierte Semantik: PP-Anbindung in einem deutschen HPSG-Fragment. In S. Mehl, A. Mertens, and M. Schulz, editors, *Präpositionalsemantik und PP-Anbindung*, number SI-16 in Schriftenreihe Informatik, pages 39–47. Gerhard-Mercator-Universität, Gesamthochschule Duisburg, 1996.
- [Wau96] Oliver Wauschkuhn. Ein Werkzeug zur partiellen syntaktischen Analyse deutscher Textkorpora. In D. Gibbon, editor, *Natural Language Processing and Speech Technology. Results of the 3rd KONVENS Conference (Bielefeld)*, pages 356–368, Berlin, 1996. Mouton de Gruyter.

Initial Results on Hybrid Connectionist Bilingual Alignment

Joseph Chen and Stefan Wermter
University of Hamburg, Computer Science Department
Vogt-Kölln-Straße 30, D-22527 Hamburg, Germany
{chen,wermter}@informatik.uni-hamburg.de

Abstract

We examine phrase alignment in a hybrid connectionist framework. We describe the architecture and the learning algorithm of our approach. Simulations have been carried out to demonstrate the feasibility of this approach using two corpora. Although the results of our approach are still at an early stage, we found that a hybrid approach to phrase alignment has the potential to provide good results using relatively little training data. While there have been a lot of statistic approaches to alignment, to the best of our knowledge this is the first hybrid connectionist approach for phrase alignment.

1 Introduction

Recently, the field of computational linguistics has seen a remarkable trend for processing restricted tasks. Rather than using traditional linguistic symbolic grammar formalisms (Chomsky 75) new learning approaches have been examined, especially for restricted tasks and domains which may not need a deep structural analysis. Among these alternative approaches are statistic (Chen 96; Brown *et al.* 93; Brown *et al.* 92; Cole *et al.* 95), connectionist (Elman 90; Chalmers 90; St. John & McClelland 90; Munro *et al.* 91) and hybrid approaches (Wermter & Weber 97; Wermter 95; Bengio *et al.* 95). These approaches have been used for learning restricted tasks like word sense disambiguation, syntactic tagging, language modeling, etc. We believe that these restricted tasks provide an appropriate framework for the applicability of a new hybrid connectionist learning approaches in natural language processing.

In this paper we describe initial experiments using a hybrid statistic/symbolic/connectionist approach for the alignment and collocation in bilingual corpora. The alignment problem is to find the corresponding words or phrases from two sentences or phrases from two different languages. An alignment component can help to compile a bilingual lexicon automatically and extract the idiomatic usage from a parallel corpus such as Canadian or Hong Kong Hansard. Furthermore, this task can be seen as a step towards more sophisticated machine translation approaches based on structured templates and a bilingual knowledge base (Kaji *et al.* 92).

Previous approaches for alignment are mostly statistic approaches (Chen 96; Gale & Church 93; Wu 94; Fung & Church 94; Fung & McKeown 94; Kay & Röscheisen 93; Xu & Tan 96). These approaches collect the statistic characteristics from large bilingual corpora and construct a model for the context without making much effort to analyze the sentences syntactically or semantically. That is, purely statistic approaches “learn” associations between lexical items (words) rather than using syntax or semantics. Therefore, purely statistic approaches need extremely large bilingual corpora so that they can estimate the probability that a certain lexical item (word) is associated with a certain lexical item in a different language. However, such large bilingual corpora of gigabyte size or more are hard to collect. Therefore, the question arises whether there are other learning approaches which use more input knowledge but much less training data.

In this paper, we use a statistic/symbolic/connectionist framework to attack the alignment problem. Our hybrid approach uses symbolically interpretable syntactic and morphological input and output representations for words in order to support a reduction of the training data compared with the large number of lexical words in other statistic approaches. Furthermore, our hybrid approach uses statistic techniques for data compression only as well as connectionist techniques for the main learning part.

For learning alignment we have to rely on manually translated training material. In general, there may be more than one satisfactory alignment and in some cases it may be done differently by different individuals. Therefore, robustness is an essential feature for alignment. Furthermore, even if some words

are neglected, it is desirable that most of the corresponding phrases are aligned as good as possible. The learnability and robustness of a connectionist framework appears to be a good candidate for a robust alignment task. On the other hand, the readily available symbolic information for the language representations can still be used. This motivates our hybrid approach to the alignment problem.

2 Alignment for Language Translation

Alignment is the problem to find corresponding parts in sentences or phrases from two languages. For example, assume the following two phrases have to be aligned.

```

DAS      _____ THE
ZEITALTER ——— TIME
DES      _____ OF
          |         THE
          |         ABSOLUTISMUS — ABSOLUTISM

```

In this example the corresponding words have been aligned by lines to demonstrate their correspondence. In general, it is possible to translate such phrases differently but we assume that a reasonable alignment is given and our goal is to try to learn and generalize this alignment as good as possible. In our example most words correspond but there is also one case where one word (DES) has been translated with two words (OF THE).

There are many other phenomena which have to be considered for language alignment and language translation. For instance, one word may not have a corresponding word in another language and therefore it has to be circumscribed. For example, from German to Chinese complicated subordinate clauses are usually avoided and translated as several simple sentences. The single compound noun in German has to be translated into several English words, possibly in a phrasal form.

Furthermore, morphological knowledge may differ a lot between two different languages. For example, the noun gender, verb conjugation and noun declension are not uniformly present in every language. Functional words such as gender or number markers may have to be introduced in such cases. Some categories may be missing in one language while they remain very important in the other language. An example is the quantitatives in Chinese/Japanese and articles in English/German/French.

Based on these observations one can clearly see that there are many difficult issues involved in alignment. Our goal is to examine whether it may be possible to *learn* to align phrases using a hybrid approach.

3 Description of the Data Material

3.1 Corpus 1: Artificially Generated L0

The first corpus is automatically generated by a context-free grammar. The grammar is based on a miniature language known as L0 (Weber & Stolcke 90; Feldman *et al.* 90). We construct a corpus of sequences of syntactic/semantic categories for a pre-specified depth within the grammar. The grammar is illustrated below.

```

s→np vp
np→det np1 | det np1 conj det np1
vp→vi pp | vt np
np1→obj | shade obj | size obj | size shade obj
pp→rel np
rel→rel1 | far rel1

```

So a sentence “s” can consist of a noun phrase “np” followed by a verb phrase “vp”. A noun phrase “np” can be either a determiner “det” followed by a certain noun phrase “np1” or two such noun groups connected by a conjunction “conj”. A verb phrase can be an intransitive verb “vi” with a prepositional phrase “pp” or a transitive verb “vt” with a noun phrase “np”. A certain noun phrase “np1” can either consist of an object “obj” or a shade object sequence or a size object sequence or a size shade object sequence. A prepositional phrase contains a relationship “rel” followed by a noun phrase “np”. A relationship “rel” is a certain relationship “rel1” optionally preceded by the word “far”.

Using this context-free grammar it is possible to automatically generate certain templates, for instance the sequence “det obj vt det shade obj”. We will use such sequences in a simple monolingual alignment

task in which every word is aligned with itself. So for this corpus the task would be to learn a word-by-word alignment based on the presentation of many aligned examples. However, the rule of a word-by-word alignment is not known prior to learning and has to be learned from the examples.

3.2 Corpus 2: Real-World Title Phrases

While our first corpus contains artificially generated sequences for alignment, the second corpus contains real-world title phrases. This title corpus contains several thousand book titles extracted from the bibliographic entries in the library (Wermter 95). We randomly chose 100 titles from this corpus and manually translated them to English. The German words of the titles are aligned with their corresponding English translations. For example, in the following title phrase and its translation we also give the numbers of the corresponding word positions.

Das Zeitalter des Absolutismus

The Time of the Absolutism

Alignment: ((1 ⇔ 1)(2 ⇔ 2)(3 ⇔ 3 4)(4 ⇔ 5))

Thus (1⇔1) means we align “Das” with “The”, and (3 ⇔ 3 4), “des” with “of the”. In fact we could say that ((1 2 3 4 ⇔ 1 2 3 4 5)) and ((1 2 ⇔ 1 2) (3 4 ⇔ 3 4 5)) are also valid alignments. However, our guideline is to provide the most detailed correspondences and therefore to divide into as many corresponding parts as possible.

Most titles are noun phrases and prepositional phrases. Their length is from one word to thirteen words. The German titles often contain long compound nouns which have to be translated into several English words or phrases. Although many of the phrases contain word-by-word correspondences there are also some translations which do not contain easy correspondences. For example, the title

Einfuehrung in die Geologie Europas

Introduction to the Geology of Europe

Alignment: ((1 ⇔ 1)(2 ⇔ 2)(3 ⇔ 3)(4 ⇔ 4)(5 ⇔ 5 6))

contains a genitive form in German “Geologie Europas” which has been translated to a prepositional phrase “Geology of Europe”.

4 A Hybrid Architecture

4.1 Overview of the Architecture

Our initial architecture is motivated by our goal of learning robust alignment. Therefore, we want to exploit the learning capabilities of connectionist networks as well as the advantageous interpretation capabilities of symbolic input/output representations. For reducing the size of the input/output representations we want to take advantage of the compression capabilities of statistic techniques. Our proposal for a symbolic/statistic/connectionist aligner is illustrated in Figure 1. The architecture is modular so that the task performed by a specific module can be replaced by another module as long as the input-output characteristics are preserved. This allows for easy changes in the overall hybrid architecture.

4.2 Symbolic Lexicon Module

First the orthographic transcripts of the source language phrases/sentences and their target language translations are sent to a symbolic lexicon module. Each word is mapped into a vector whose units contain its possible morphologic and grammatical features. In the L0 miniature corpus, the representation of each symbol is listed below.

conj	[0 1 0 0 0 0 0 0 0 0 0 0]
det	[0 0 1 0 0 0 0 0 0 0 0 0]
dist	[0 0 0 1 0 0 0 0 0 0 0 0]
obj	[0 0 0 0 0 1 0 0 0 0 0 0]
rell	[0 0 0 0 0 0 0 1 0 0 0 0]
shade	[0 0 0 0 0 0 0 0 1 0 0 0]
size	[0 0 0 0 0 0 0 0 0 1 0 0]
vi	[0 0 0 0 0 0 0 0 0 0 0 1]
vt	[0 0 0 0 0 0 0 0 0 0 0 1]

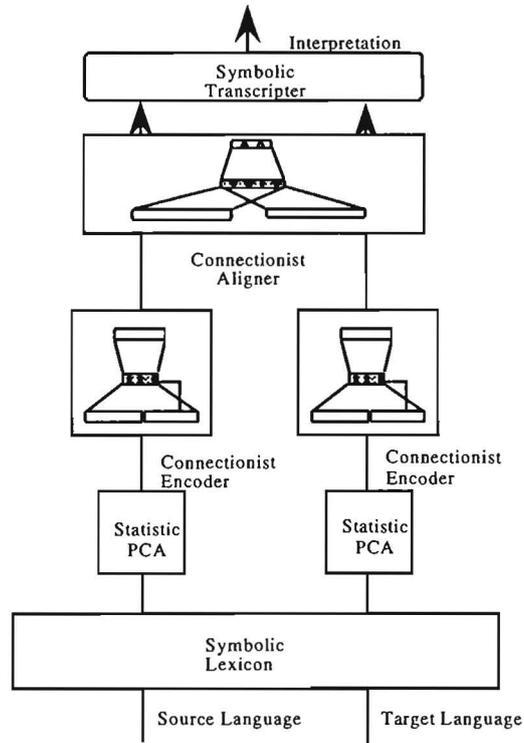


Figure 1: Modular design of a bilingual aligner

In the title corpus a German word was represented as a 22-dimensional activation vector. Each unit in the activation vector represents a linguistic feature according to Table 1. These features were chosen as an initial set of important underlying knowledge for alignment, and they are based on categories from the Langenscheidt dictionary. An English word was encoded as a 16-dimensional activation vector according to Table 1. The English and German representations differ slightly due to differences with respect to declensions and conjugations. The vocabulary of the German phrases comprises 294 different words, and that of the English phrases 332 words. The difference is due mostly to compound nouns which have to be translated as longer phrases in English.

4.3 Statistic Preprocessing: Principal Component Analysis

Recurrent neural networks have relatively expensive numerical computations due to the recurrent connections. So it is an important issue to keep the size of a network small in order to support fast training. One

German Lexicon		English Lexicon	
Part-of-Speech	Morphologic/ Functional	Part-of-Speech	Morphologic/ Functional
num (numeral)	time (time modifier)	num (numeral)	time (time modifier)
n (noun)	m (masculine)	n (noun)	pl (plural)
pn (proper noun)	f (feminine)	pn (pronoun)	gen (genitive)
adj (adjective)	nt (neutral)	adj (adjective)	itrans (intransitive)
prep (preposition)	pl (plural)	prep (preposition)	trans (transitive)
det (determinate article)	dat (dative)	det (determinate article)	ord (ordinal)
adv (adverb)	acc (accusative)	adv (adverb)	
v (verb)	gen (genitive)	v (verb)	
art (indeterminate article)	trans (transitive)	pron (pronoun)	
pron (pronoun)	itrans (intransitive)	conj (conjunction)	
conj (conjunction)	ord (ordinal)		

Table 1: Features in the lexicons.

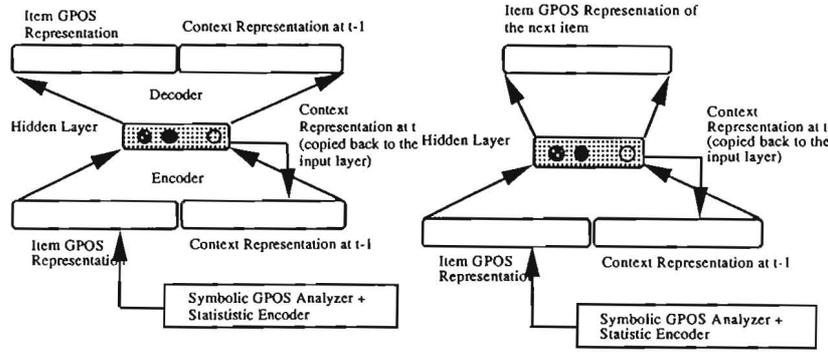


Figure 2: RAAM and SRN encoder architectures

possibility to compress the local input representation of a single word is to apply principal component analysis (PCA) (Gonzalez & Wintz 77) to the input data. Then most important leading components of the transformed vector can be chosen as the representative of the word representation (in our case 5).

4.4 Encoding Symbolic Sequences in Connectionist Networks

4.4.1 Context Representation Modules

The central modules in this architecture are the artificial neural networks (ANNs) that learn the crucial encoding tasks in alignment. We examine two types of networks for sequential processing within the overall hybrid architecture: a recurrent auto-associative memory (RAAM) (Pollack 90) and a simple recurrent network (SRN) (Elman 90).

In the RAAM version of the architecture, there are two ANN encoders and decoders, one pair for each language (see also Figure 1 and Figure 2 left). Each encoder and decoder is trained to auto-associate the words in each phrase/utterance. In this study, we coarsely divide the word representation into two locally encoded feature parts: a major part of speech code and a minor morphologic/functional code. Both are referred to as generalized part of speech representation (abbreviated as GPOS). The two ANN are working independently and can develop different compressions of phrases or sentences according to the specific language. The RAAM architecture is illustrated in Figure 2 left.

In the second version of our architecture we use simple recurrent networks (Elman 90) as an alternative for encoding and decoding the contextual information in a phrase (see Figure 2 right). The contextual information softly degrades to the left of the current word, which makes this model more plausible than a fixed n-gram model. More specifically, the activation in the hidden layer encodes the left context from the beginning of the phrase to the current word. While the RAAM networks are used as autoassociators of a sequence of words these SRN are used to predict the representation of the next word in a sequence. The last word in the phrase should predict a particular symbol for the end of the phrase which is represented by the average activation of each unit. The context layer is reset between two different phrases.

4.4.2 Connectionist Learning Algorithm: Conjugate Gradient

In our implementation, the neural networks are trained with the conjugate gradient method (Press *et al.* 92). The backpropagation method (Rumelhart *et al.* 86) is a special gradient descent method which is widely known in the unconstrained optimization theory (Alpsan *et al.* 95; Kalman & Kwasny 94). While applying backpropagation for recurrent networks, especially in the case of RAAM, the choice of learning parameters is usually a crucial issue. The learning rate, for example, cannot be chosen too big since the hidden layer representations have to be developed as well. However, too small learning rates also lead to slow learning.

Inspired by the arguments by Kalman (Kalman & Kwasny 94), we write the error function f as,

$$f(v) = \sum_p \sum_n \frac{(o_n - t_n)^2}{1 - o_n^2}$$

where o is the activation in the output layer, n is the position of a unit, p is the number of patterns, v is the weight vector, and t is the training target. Our activation is restricted in $(-1,1)$. A hyperbolic

tangent threshold function is used for each unit.

4.5 FF-Networks for Aligning RAAM and SRN Representations

So far we have described the overall hybrid architecture, a statistic preprocessor for dimension reduction and two alternative versions to encode partial sequences of words as RAAMs and SRNs. We will now focus on the issue of learning the alignment itself. The alignment is considered to be a binary relation which is bidirectional. We use a feed forward network architecture as illustrated in Figure 3.

The word representation and the context representation, which is the activation in the hidden layer of either SRN or RAAM encoder, are concatenated as the input vector. In the SRN implementation, the context representation is the hidden activation at the current word in a phrase, and the word representation is the feature representation. In the RAAM implementation, the context representation is the final hidden activation. The word representation, however, is the hidden activation at the current word. This schema is chosen because the RAAM can decode the word itself given a hidden activation. The concatenation is denoted in Figure 3 as “Word+Context Representation”.

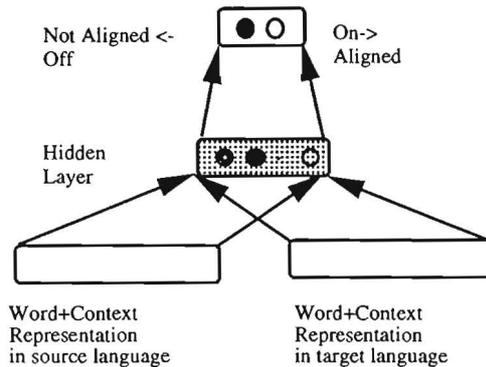
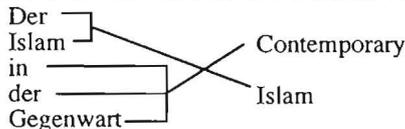


Figure 3: Feedforward network as an aligner

5 Experiments with Various Hybrid Architectures

5.1 Measurement of Performance

Our measurement of correctness is based on the number of individual alignments. The individual alignment is a set of word pairs and can be denoted by a pair of integers. Each integer represents the index number of the word in the German and English phrase, respectively. Consider the following example:



There are $5 \times 2 = 10$ individual alignments, each of which can be either positive or negative. In the example above, the individual alignment (Der \Leftrightarrow Islam) (or $(1 \Leftrightarrow 2)$) and (Islam \Leftrightarrow Islam) (or $(2 \Leftrightarrow 2)$) are positive; the individual alignment (Der \Leftrightarrow Contemporary) (or $(1 \Leftrightarrow 1)$) is negative. If an individual alignment from the output of the network is positive (or negative) as indicated by its label, this outcome is counted as correct, otherwise incorrect.

The correctness is calculated by dividing the number of correct individual alignments by the total number of individual alignments. A pair is considered as “positive” if and only if the activation of the aligned-neuron is higher than that of the not-aligned-neuron. The ratio between the negative alignment and positive alignment pairs depends on the length of the phrase/sentence and the complexity of the language. We found that this ratio in the title phrase corpus is about 4. We correct this unbalance by using a normalization mechanism in preparing the training set. We modified the error function so that it can scale the error contributed by the positive and the negative alignments to the same degree.

5.2 Corpus 1: Artificial L0 Language

In our first and simplest experiment, the training corpus contains the 16 sentences generated by the L0 grammar with the maximal depth 5. The test set is generated automatically with the maximal depth 6 and it contains 144 sentences. Although the L0 corpus is a monolingual corpus, we use two independent context encoding networks and trained the networks independently. The number of hidden units in the encoding SRN is 10. There are 8 hidden neurons in the feedforward alignment network. On the training set, an alignment accuracy of 97% is achieved. The overall accuracy for the test set is 93.3% (15368/16480).

The training and test set is the same as for the SRN context encoding network. With the training set an alignment accuracy rate 100%(800/800) is achieved, in which all the positive and negative alignments are strictly positive or negative. On the test set, the alignment accuracy rate is 89.1% (14687/16480).

5.3 Corpus 2: Real World Phrases

There are 80 phrases in the training set and 20 phrases in the test set. The number of hidden units is 24 in both SRN context encoding and feed forward alignment network. For the training set, the alignment accuracy rate is 99.96% (2438/2439 words). The test set contains 20 phrases that were not in the training set. The overall alignment accuracy rate is 81.4% (367/451 words).

The network geometry is the same as in the SRN configuration. The alignment accuracy for the training set is 100% (2439/2439). On the test set, an accuracy of 84.7% (382/451) is achieved.

5.4 Examples

In this section, we will take a closer look at the individual examples from the unknown test set of the title corpus. All the following examples are chosen from the test set of the book title corpus. The architecture and architectural parameters used are those in the SRN encoder reported above.

Example 1:

ARCHITEKTUR — ARCHITECTURE
HEUTE — TODAY

The alignment feedforward network can identify all the positive and negative alignments correctly. This seems to be a trivial one-to-one case. However, one has to consider that the network does not have hard-coded structural knowledge as we noted above. Besides, this alignment pattern is not found in the training set according to our symbolic feature representation. The network is able to generalize from similar patterns which it has seen before and reaches the correct alignment.

Example 2:

MARKUS-EVANGELIUM — { GOSPEL
 OF
 MARK

This example is a one-to-three alignment. The network can establish the correct alignments. The exact pattern representation does not show up in the training set either. A similar example in the training set is “ROEMERBRIEF = LETTER TO ROMANS”, although this example has a slightly different representation. This demonstrates the generalization ability of the network.

Example 3:

EINFUEHRUNG — INTRODUCTION
 IN — TO
 DIE — THE
TEXTVERARBEITUNG — WORD
 PROCESSING

The network established all the correct alignments, even though the exact pattern representation was not in the training set. A similar pattern in the training set is “EINFUEHRUNG IN DIE MUSIKALISCHE VOLKSKUNDE = INTRODUCTION TO THE MUSICAL ETHNOLOGY”, which is a one-to-one alignment. There are other examples in the training set which are one-to-two noun phrasal patterns. The network seems to combine these two regularities to accomplish the alignment shown here.

Example 4:

FERTIGUNGSGERECHTE } — PRODUCTION
 LAW
 OF
III-V-EPITAXIE } — III-V-EPITAXY

Finally we give an example for the robustness of the network using an intentionally very weak translation. Instead of translating a certain concept “FERTIGUNGSGERECHTE III-V-EPITAXIE” as “suitable III-V-EPITAXY” or “appropriate III-V-EPITAXY” we used a weak translation of FERTIGUNGSGERECHTE III-V-EPITAXIE as PRODUCTION LAWS OF III-V-EPITAXY. These patterns were very different from similar examples in the training set. However, the network can still output a “useful” result. That is, all the words in the German phrase are aligned with all the words in the English phrase. This is interesting phenomenon since a network did not assign all the alignment pairs as negative.

In summary what we have accomplished so far is a learning mechanism for easy forms of alignment. That is, for smaller phrases the alignment is quite reliable and robust as we have demonstrated above. More work has to be done if we want to align very long phrases or sentences since the structure and semantics of the corresponding words and subphrases may be very different. All the examples of the training set had been learned perfectly. Most of the remaining errors in the unknown test set show up in very long phrases. However, at the current stage, our approach can learn and generalize the alignment of smaller phrase groups quite reliably already.

6 Discussion and Conclusion

We have proposed an hybrid approach for alignment. While there have been a lot of statistic approaches to alignment, to the best of our knowledge this is the first hybrid connectionist approach for alignment. This research is still at an early stage compared to many traditional alignment approaches which use mainly statistic techniques. However, the purely statistic approaches suffer from the large quantities of training material. In our approach, we try to address alignment using smaller corpora since gigabyte or terabyte size corpora are not readily available in all domains.

A symbolic lexicon component allows to associate category knowledge with lexical words. Since category representations of words can grow large we use a statistic preprocessing step to reduce the size of the input representations. Finally, all the actual sequence encoding and alignment is learned using various connectionist networks. In contrast to symbolically and manually encoded aligners connectionist networks allow the learning of alignment regularities from examples and therefore should be more robust. On the other hand, the use of symbolic category representations allows to test alignment strategies for small and medium corpora, which are areas for which statistic alignment on lexical word items does not perform well.

We have used and compared RAAM and SRN networks for encoding sequences of words. The autoassociation task in RAAMs and the prediction task in SRN were not particularly developed for the alignment task. Furthermore, the performance of these individual networks for autoassociation and prediction was not perfect. In general, this means that the encoding networks did not yet contain much domain or task knowledge. Given this restricted autoassociative or predictive context knowledge the performance of the alignment network is quite reasonable already.

This leads us to expect that this approach has some potential, especially since we still did not yet make use of important underlying knowledge for alignment (e.g. semantics), since the encoding networks did not yet use alignment specific knowledge, and since initial results with simple recurrent networks on a real-world corpus were quite encouraging. We expect that a suitable hybrid connectionist architecture which is able to systematically encode a natural language sequence with more semantics could also improve the performance scaling up to bigger phrases and sentences.

References

- (Alpsan *et al.* 95) Dogan Alpsan, M. Towsey, O. Ozdamar, A.C. Tsoi, and D.N. Ghista. *Efficacy of Modified Backpropagation and Optimization Methods on a Real-World Medical Problem*. *Neural Networks*, 8(6):945–962, 1995.
- (Bengio *et al.* 95) Yoshua Bengio, Yann LeCun, Craig Nohl, and Chris Burges. *LeRec: A NN/HMM Hybrid for On-Line Handwriting Recognition*. *Neural Networks*, pages 1289–1303, 1995.
- (Brown *et al.* 92) Peter F. Brown, V.J. Della Pietra, P.V. de Souza, J.C. Lai, and R.L. Mercer. *Class-Based n-gram Models of Natural Language*. *Computer Linguistics*, 18(4):467–479, 1992.

- (Brown *et al.* 93) P.F. Brown, S.A. Della Pietra, V.J. Della Pietra, and R.L. Mercer. *The Mathematics of Statistical Machine Translation: Parameter Estimation*. *Computational Linguistics*, 19(2):263–311, 1993.
- (Chalmers 90) D.J. Chalmers. *Syntactic Transformations on Distributed Representations*. *Connection Science*, 2(1&2):53–62, 1990.
- (Chen 96) Stanley F. Chen. *Building Probabilistic Models for Natural Language*. PhD Thesis, Harvard University, 1996.
- (Chomsky 75) Noam Chomsky. *The Logical Structure of Linguistic Theory*. The University of Chicago Press, Chicago, USA, 1975.
- (Cole *et al.* 95) Ronald A. Cole, J. Mariani, H. Uszkoreit, A. Zaenen, and V. Zue. *Survey of the State of the Art in Human Language Technology*. Survey, National Science Foundation, Directorate XIII-E of the Commission of the European Communities, Center for Spoken Language Understanding, Oregon Graduate Institute, November 21 1995.
- (Elman 90) Jeffery Elman. *Finding Structure in Time*. *Cognitive Science*, 14:179–211, 1990.
- (Feldman *et al.* 90) Jerome A. Feldman, G. Lakoff, A. Stolcke, and S.H. Weber. *Miniature Language Acquisition: A touchstone for cognitive science*. Technical Report TR-90-009, International Computer Science Institute, 1990.
- (Fung & Church 94) Pascale Fung and K. W. Church. *K-vec: A New Approach for Aligning Parallel Texts*. Technical Report cmp-lg/9407021, CS Dept, Columbia University, N.Y., AT&T Bell Laboratories, 1994.
- (Fung & McKeown 94) Pascale Fung and K. McKeown. *Aligning Noisy Parallel Corpora Across Language Groups: Word Pair Feature Matching by Dynamic Time Warping*. Technical Report cmp-lg/9409011, CS Dept, Columbia University, N.Y., 1994.
- (Gale & Church 93) William A. Gale and K.W. Church. *A Program for Aligning Sentences in Bilingual Corpora*. *Computational Linguistics*, 19(1):75–102, 1993.
- (Gonzalez & Wintz 77) Rafael C. Gonzalez and P. Wintz. *Digital Image Processing*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1977.
- (Kaji *et al.* 92) Hiroyuki Kaji, Y. Kida, and Y. Morimoto. *Learning Translation Templates from Bilingual Text*. In *COLING-92*, volume 2, pages 672–678, Nantes, 1992.
- (Kalman & Kwasny 94) Barry Kalman and Stan C. Kwasny. *High Performance Training of Feedforward & Simple Recurrent Network*. Technical Report WUCS-94-29, Washington University, 1994.
- (Kay & Röscheisen 93) Martin Kay and Martin Röscheisen. *Text-Translation Alignment*. *Computational Linguistics*, 19(1):121–142, 1993.
- (Munro *et al.* 91) Paul Munro, C. Cosic, and M. Tabasko. *A Network for Encoding, Decoding and Translating Locative Prepositions*. *Connection Science*, 3(3):225–240, 1991.
- (Pollack 90) Jordan B. Pollack. *Recursive Distributed Representations*. *Artificial Intelligence*, 46:77–105, 1990.
- (Press *et al.* 92) William H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Minimization or Maximization of Functions*. In *Numerical Recipes in C*. Cambridge University Press, 1992.
- (Rumelhart *et al.* 86) David E. Rumelhart, G.E. Hinton, and R.J. Williams. *Learning Internal Representations by Error Propagation*. In J. L. David E. Rumelhart, McClelland, editor, *Parallel Distributed Processing*, volume 1, Foundation, pages 318–362. MIT Press, 1986.
- (St. John & McClelland 90) Mark F. St. John and J. L. McClelland. *Learning and Applying Contextual Constraints in Sentence Comprehension*. *Artificial Intelligence*, 46:217–257, 1990.
- (Weber & Stolcke 90) Susan H. Weber and A. Stolcke. *L0: A Testbed for Miniature Language Acquisition*. Technical Report TR-90-010, International Computer Science Institute, July 1990.

- (Wermter & Weber 97) Stefan Wermter and V. Weber. *SCREEN: Learning a Flat Syntactic and Semantic Spoken Language Analysis Using Artificial Neural Networks*. *Journal of Artificial Intelligence Research*, 6(1):35–85, 1997.
- (Wermter 95) Stefan Wermter. *Hybrid Connectionist Natural Language Processing*, volume 7 of *Chapman & Hall Neural Computing Series*. Chapman & Hall, London, 1995.
- (Wu 94) Dekai Wu. *Aligning a Parallel English-Chinese Corpus Statistically with Lexical Criteria*. Technical Report cmp-lg/9406007, Dept. of Computer Science, University of Science & Technology, Clear Water Bay, Hong Kong, 1994.
- (Xu & Tan 96) Donghua Xu and C. L. Tan. *Automatic Alignment of English-Chinese Bilingual Texts of CNS News*. Technical Report cmp-lg/9608017, Department of Information system and Computer Science, National Univerisity of Singapore, Singapore 119260, 1996.

Identifying dialogue acts: linguistics or statistics?

Manfred Stede

Technische Universität Berlin

Projekt KIT-VM11 Sekr. FR 6-10

Franklinstr. 28/29

D-10587 Berlin

Email: stede@cs.tu-berlin.de

Abstract

At some point in their processing, natural language dialogue systems typically perform some pragmatic analysis to compute the *dialogue act* (or *speech act*, *dialogue move*, *illocution*, etc.) underlying an utterance. The VERBMOBIL system, despite the fact that it is not engaged in the dialogue but only mediates it, is no exception. At present, both symbolic reasoning and statistical recognition are used for this purpose. This paper discusses the relative merits of both approaches and of various ways of integrating the two, which are currently subject to experimentation.

1 Dialogue acts in VERBMOBIL

Natural language dialogue systems typically analyze utterances on various levels of abstraction, and often the “deepest” representation they construct involves *dialogue acts* (DAs). These are constructs that abstract from specific lexical, syntactic, and semantic information, and represent only the pragmatic force that is relevant for the processing system. In the VERBMOBIL project [Wahlster 1993], DAs are organized in a hierarchy reflecting their specificity, and — to a certain extent — domain-dependency: the most specific DAs are geared towards the task of appointment scheduling. Figure 1 shows the hierarchy of DAs used in phase I of the project [Jekat et al. 1995]. At present, the list is being revised for phase II; nonetheless, we can base the following discussion on the hierarchy shown in the figure.

Since VERBMOBIL employs several lines of analysis, varying on the scale from “deep” to “shallow”, different methods for recognizing the DA of an utterance are in use. Specifically, within the DIAKON component (dialogue processing and contextual evaluation) both statistical and rule-based recognition schemes are at work. The following two sections briefly describe these schemes. Then, section four examines the relative merits of the competing approaches and discusses possibilities for integrating them into a single framework.

2 Recognizing dialogue acts, linguistically

In VERBMOBIL, the contextual evaluation (ConEval) module is given the syntactic/semantic representation of an utterance and asked to perform certain disambiguations and to compute

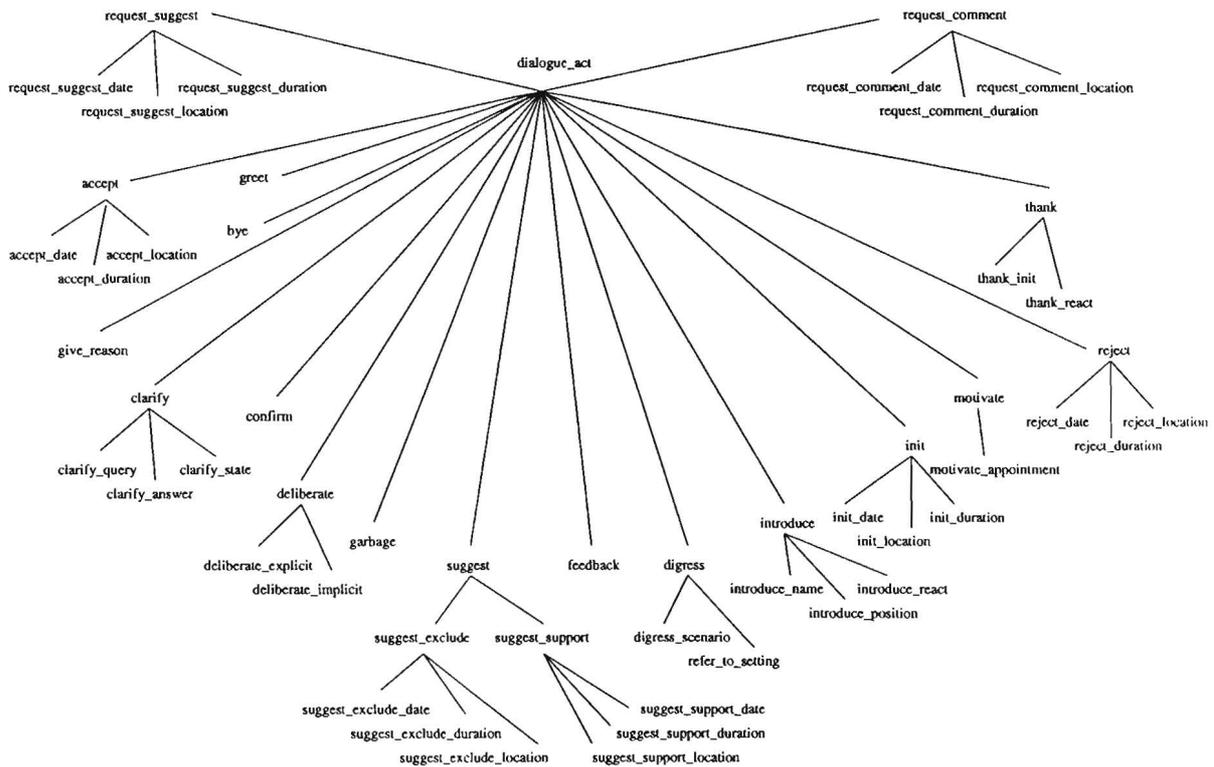


Figure 1: Dialogue acts used in Verbmobil-I

the underlying dialogue act. To this end, ConEval constructs an utterance representation in FLEX [Quantz et al. 1996], a description logic (DL) language, which is the basis for its calculations.

To identify the dialogue act, we use an extension to the standard description logic: an apparatus of *weighted default rules*. The mechanism is described in [Schmitz, Quantz 1995]. Briefly, these rules check for the presence of certain features and infer a *suggested* dialogue act; every rule is associated with a numerical weight, and after processing the rules, the accumulated weights decide on the most likely DA, which is then added to the representation of the utterance.

For the rules, the primary source of information is the verb; we have defined semantic verb classes such as PROPOSING-VERB and SUITING-VERB, and membership is tested on the left-hand side of many recognition rules. (In practice, of course, information on the verb class is combined with other information first, most prominently with negations.) Other features used in the rules pertain to sentence mood, syntactic environment, lexical collocations, semantic classes of arguments and adjuncts, the preceding DA, and the DA-prediction, to which we will turn in the next section.

For illustration, here is one of the rules used to recognize the DA REQUEST-COMMENT-DATE. It checks whether

- the conceptual content (which has already been computed from the verb and its modifiers) is SUITING,

- the direct object of the verb is a pronoun referring to the hearer,
- the utterance is a question, and
- the previous DA was SUGGEST-SUPPORT-DATE.

If all these features are present, the weight 90 is added to the DA REQUEST-COMMENT-DATE.

```
conc:suiting and
the(arg2,pron_ref:he) and
sent_mood:ynq and
the(direct_pred,suggest_support_date)
~90~> request_comment_date
```

Dialogue act identification interacts heavily with other aspects of ConEval’s work, especially with disambiguation and anaphora resolution. Consider the utterance *Ja, das wuerde passen*. Here, the verb *passen* is ambiguous — most likely in our domain, the reading *to suit* is intended. Then, a translation such as *Yes, that’s suitable* or *Yes, that would work* can be produced. To determine the precise DA, however, the referent of the anaphoric pronoun *das* needs to be determined. When it has been resolved, the general illocution ACCEPT can be further specified to ACCEPT-DATE or ACCEPT-LOCATION. In other cases, anaphora resolution also affects the translation directly.

Sometimes, computing the DA involves inferencing; in the VERBMOBIL domain this is in particular the comparison of temporal expressions. For instance, the DA CONFIRM captures utterances that wrap up the negotiation and possibly re-state the outcome, i.e. the date the partners agreed on. An utterance like *Treffen wir uns am dritten Mai* can be a SUGGEST, or a CONFIRM, if May 3rd has been negotiated earlier in the dialogue. Inferences of this kind are part of ConEval’s operation and contribute to proper DA identification.

Unfortunately, evaluating the performance of the recognition rules is not easy, because we rely on fully analyzed utterances that have passed through VERBMOBIL, and for which we have constructed the DL representation. In contrast to the “raw” data, or dialogue transliterations, which are used by the statistical approaches described below, input for our module is not available in large quantities. With limited evaluation possibilities, we determined a rate of close to 80% of correctly recognized DAs.

3 Predicting and recognizing dialogue acts, statistically

In contrast to the symbolic approach just described, the dialogue-processing group of VERBMOBIL has developed statistical methods to predict the next dialogue act on the basis of the dialogue history [Reithinger et al. 1996]. The basic idea is to exploit the analogy between predicting the next act in a dialogue on the one hand, and predicting the next word in a sentence. Accordingly, Reithinger et al. adopt methods from speech recognition (n-grams, Markov chains) to the task of DA prediction. Using only the 18 general DAs from figure 1, they report a hit rate of 40% when predicting the single next DA; when predicting a set of 2 DAs, the correct one is in this set in 65% of all cases; and with a set of 3, this rate goes up to 75%. When all 42 DAs are used for the task, however, the performance drops by 10%.

Recently, Reithinger and Klesen [1997] developed a DA classifier that tries to identify the DA for a given input string, using a linear interpolation of uni- and bigrams. They evaluated the scheme using the full set of 43 DAs. As is to be expected, the results are best for DAs that are typically realized by highly conventionalized formulas (such as GREET, THANK, BYE): both precision and recall are close to or better than 90%. For the most frequently occurring DAs, which are important for dialogue development, recall rates range from about 40% to 80%, and precision is between 40% and 70%. For the remaining DAs, results are rather bad due to the small number of occurrences in the training data. Somewhat surprisingly, providing the classifier with the additional information of the dialogue history does not improve the recall rate very much. The authors note that the additional knowledge yields an increase of “up to 3%”. — Finally, the same procedures were run with the reduced set of 18 general DAs, which lead to an overall increase of the rates by roughly 2%.

Statistical DA recognition is also being done in the “shallow processing” mode of VERBMOBIL. Shallow processing operates when no translation is needed, and the analysis is carried out solely for the purpose of keeping track with the dialogue. Also on the basis of the bare input string without any syntactic or semantic information, Mast et al. [1996] identify dialogue acts with classification trees and report recognition rates of 46% when using 16 DAs. With an alternative approach of n-grams, for n=5 the success rate is 69%, and — interestingly — for n=2, it is almost the same: 67%.

4 Symbols, or numbers, or both?

4.1 Strengths and Weaknesses

The obvious advantages of statistical methods are their efficiency and flexibility. Once the training procedure is set up, and the data is at hand, automatization takes over. Furthermore, at run-time the statistical approach produces answers quickly.

On the other hand, the reliance on a sufficient amount of training data turns into a disadvantage when it comes to dealing with DAs that occur relatively infrequently. The tables given by Reithinger and Klesen [1997] show that only 8 DAs achieve recall rates of more than 70%, and 16 DAs of more than 40%. Among the ones with low recall are some that are quite important for dialogue progression, such as CONFIRM (33%), ACCEPT-LOCATION (26%), or SUGGEST-SUPPORT-DURATION (22%). To a good extent, these problems are due to a lack of enough training data containing these DAs. This, however, is not the only reason, as argued below.

As for exploiting the dialogue history with approaches from language modelling, it does not seem self-evident that the analogy between speech recognition and DA recognition really holds. Sequences of words in a sentence follow grammatical rules and thus are relatively predictable; exchanging moves in a dialogue also follows certain patterns, but these are less strict. In the current VERBMOBIL domain, we are dealing with highly stereotyped, goal-directed dialogues, which seem to be quite amenable to statistical modelling, but with any less-restricted interaction, the situation is different.

Regarding the rule-based approach, its “traditional” disadvantage is the fact that rules are hand-crafted, and weights are adjusted manually. In the ConEval group, experiments to

induce decision trees automatically and to determine weights using the Simplex method for linear equations are under way, but there are no results yet. As pointed out above, ConEval has to work with fully analyzed utterances in the DL representation, so we cannot perform any training on raw strings. In addition, for an online interpreting system, a drawback of knowledge-based analyses is that they take time.

On the conceptual side, rules have difficulties with those DAs that do not give good clues from the verclass etc., such as GIVE-REASON, where — in principle — a basically unrestricted range of propositions is to be expected. Here, ConEval uses the preceding DA as an important source of information: After what kind of utterance is it reasonable to give a reason? This is usually a good clue, and it seems that statistical recognition, looking back into the dialogue history, cannot do much better in these cases. Recall the result from Mast et al., who found that a bigram approach performs almost just as well as a pentagram approach — and the relevant bigrams are encoded in our symbolic rules just as well.

The central advantage of a knowledge-based approach is its ability to perform the necessary linguistic analysis and the reasoning that is required in many cases. Recall the example of comparing temporal expressions mentioned above. Similarly, the difference between SUGGEST-SUPPORT-DATE and SUGGEST-EXCLUDE-DATE is an important one, but often it can only be determined by analyzing the scope relations between verbs, negations, and various adverbs. As an indication for these problems, Reithinger and Klesen's [1997] statistical method achieves a precision rate of just 67% for SUGGEST-SUPPORT-DATE, which is not quite satisfactory for a DA of such critical importance in the domain.

Also, DA identification often interacts with disambiguation (and that is why ConEval handles both tasks). To give but one example, an utterance like *Das wäre bei mir am besten* could accept a date just proposed by the dialogue partner, or suggest a location for a meeting, depending on the referent of *das*. Fine-grained distinctions can be quite important and require fine-grained analysis. The statistical methods, however, prefer to work with as few DAs as possible (and thus with relatively abstract ones), which for them renders the whole task easier. By focusing on 16 or 18 abstract DAs, distinctions that can be relevant for translation are lost.

In conclusion, we see a symbolic reasoning component as indispensable for the general problem of reliable DA identification. Then, is there a good way of adding to the reasoning component the advantages delivered by a statistical recognition component?

4.2 Integration

To determine the prospects for integrating symbolic and statistical DA recognition, let us consider the question in what situations the statistical result would be most helpful, from the perspective of the ConEval module. Here, four possibilities are theoretically conceivable.

- **History** The DAs *beyond* the immediate predecessor are important. In this case, SR would be “at its best”. However, as already noted, experience indicates that this situation is extremely rare in the VERBMOBIL dialogues.
- **Accuracy** If the recognition results delivered by statistical analysis were — on average — overall better than those of ConEval, it would be sensible to rely solely on statistics.

This, however, is not the case.

- **Robustness** In cases where ConEval’s rules fail, SR can come to rescue. This sounds very good but is problematic in practice. First, it presupposes that ConEval can actually *notice* a failure. To a certain extent, this is possible, when the computed DA is GARBAGE. But, second, we need to know the *reason* for the failure: Either the recognition rules are imperfect, or the input is in fact bad. (The latter possibility again splits in two: the syntactic/semantic analysis has failed, or the speaker’s utterance was incoherent to begin with. In VERBMOBIL, both cases have been witnessed.) In other words, the DA GARBAGE is overloaded. And third, it is not clear at all that the cases where ConEval’s rules produce a wrong result are cases where SR produces the correct result. Quite often, the problems appear to be the same for both approaches.
- **Speed** As pointed out above, ConEval might take some time, while SR is fast.

This last aspect is the one that deserves the most attention. VERBMOBIL operates in real-time, hence efficiency is a central concern. The system architecture is organized such that ConEval receives the statistical recognition result from the Dialogue module, and is in charge of delivering the dialogue act information to the transfer module upon request. In other words, whenever transfer needs the dialogue act, its progress hinges upon a response from ConEval.

Specifically, this means that when the system is confronted with a simple utterance, understanding and translating should be performed quickly, without waiting for the results of a deep analysis. If, on the other hand, the utterance is complex, then a deep analysis might very well be necessary.

Therefore, we should investigate integration possibilities geared towards the central goal of improving overall efficiency. A range of possible strategies comes to mind.

“**Glance’n’decide**” Is there were a way to “take a quick look” at an utterance and decide whether a deep analysis and reasoning will be required, or whether shallow methods will be sufficient? This would be ideal, but it is not clear how it can be done effectively.

“**First-come-first-serve**” Use the statistics-result as long as ConEval has not finished its computations, so that other modules, if they are waiting for the DA, can carry on their work. Once ConEval has produced its own result, use that. This strategy is risky, because the results might be incompatible, and the information-state of the overall system gets complicated to control.

“**High-or-low**” If ConEval’s default rules produce a result whose weight is below a certain threshold, then rather use the DA given by statistics. This method presupposes that weights indicate not only relative victories over competing DAs but also absolute confidence in the result. Fine-tuning rules, weights, and the threshold for this purpose is very difficult, though.

“**Merge-it-all**” Factor the statistics-result into the default rules right away, i.e., give additional weight to the DA delivered by statistics. Importantly, we can use different add-on weights for different DAs; in statistical recognition, the DAs have different a priori success rates, which are known and can thus be translated into our add-on weights. Also, this method could be combined with “first-come-first-serve”: Use the statistics-DA as the “quick” result in case another VERBMOBIL module needs a DA quickly, and then factor it into the default

rule computation. If the rules come up with a different result, the difference can be more or less dramatic: some DAs are similar, others are not. If the difference is dramatic, and the SR-DA has already been posted throughout the system, some mechanism needs to be found to revise it.

“Check-it-out” Take the result from statistics and decide on further activities on the basis of this result. We know a priori that some DAs have excellent recall and precision rates, and for these, ConEval need not really take any further action: GREET, BYE, THANK, INTRODUCE. Not surprisingly, these are DAs that are typically realized by stereotypical formulae, and thus ConEval, too, would have little difficulty in recognizing them. Still, there is some gain in not touching the syntactic/semantic representation of such utterances at all. For the remaining DAs, it seems reasonable to exploit the finding that statistical recognition is better with abstract DAs, possibly with only the basic illocution. We can use the statistical approach to produce this rough description quickly, and then ConEval can perform further analyses to determine the more specific DA. Again, two strategies are possible: Either use the “standard” set of default rules, or take confidence in the statistical result and invoke only a smaller set of rules dedicated to the specific illocution that has already been determined; in other words, do not even consider DAs that would be incompatible with the abstract one determined by statistics. Again, the amount of “trust” given to the statistics-result can be made dependent on the a priori success rates known for the various DAs.

Finding the most appropriate integration method is subject to experimentation in Verbmobil-II. At present, “check-it-out” seems most promising to us, and we will explore it in depth.

References

- [Jekat et al. 1995] S. Jekat, A. Klein, E. Maier, I. Maleck, M. Mast, J. Quantz. “Dialogue Acts in Verbmobil.” Universität Hamburg, DFKI Saarbrücken, Universität Erlangen, TU Berlin; Verbmobil-Report 65.
- [Mast et al. 1996] M. Mast, R. Kompe, S. Harbeck, A. Kiessling, H. Niemann, E. Nöth, E.G. Schukat-Talamazzini, V. Warnke. “Dialog act classification with the help of prosody.” In: Proceedings of ICSLP 1996, Philadelphia/PA, 1996.
- [Quantz et al. 1996] J.J. Quantz, G. Dunker, F. Bergmann, I. Kellner. *The FLEX System*. KIT Report 124, FB Informatik, TU Berlin, 1996.
- [Reithinger et al. 1996] N. Reithinger, R. Engel, M. Kipp, M. Klesen. “Predicting Dialogue Acts for a Speech-To-Speech Translation System.” In: Proceedings of ICSLP 1996, Philadelphia/PA, 1996.
- [Reithinger, Klesen 1997] N. Reithinger, M. Klesen. “Dialogue act classification using language models.” In: Proceedings of Eurospeech 97, Rhodes/Greece, 1997.
- [Schmitz, Quantz 1995] B. Schmitz and J.J. Quantz. *Dialogue-Act Type Recognition in Automatic Dialogue Interpreting*. In: Proceedings of the The Sixth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI), Leuven, 1995.
- [Wahlster 1993] W. Wahlster. *Verbmobil: Translation of face-to-face dialogues*. In: Proceedings of the Third European Conference on Speech Communication and Technology, Berlin, 1993.

DIA-MOLE: Learning Spoken Dialogue Models By Integrating Multiple Knowledge Sources

Jens-Uwe Möller

Natural Language Systems Division,
Dept. of Computer Science, Univ. of Hamburg
Vogt-Koelln-Str. 30, D-22527 Hamburg, Germany
Phone: ++49 40 5494 - 2516 / Fax: ++49 40 5494 - 2515
<http://www.informatik.uni-hamburg.de/NATS/staff/moeller.html>
<mailto:jum@informatik.uni-hamburg.de>

1 Abstract

The Dialogue Model Learning Environment supports an engineering-oriented approach towards dialogue modelling for a spoken-language interface. In difference to many other approaches a set of dialogue acts is not predefined by any theory or manually described during the engineering process, but is learned integrating different kind of data that are available in an avised spoken dialogue system. The architecture is outlined and the approach is applied to the domain of appointment scheduling.

2 Introduction

To build dialogue models for spoken dialogue systems based on human-to-human dialogues is a tremendous work: data acquisition (recording and transcribing), analysis (according to some dialogue structuring theory) and the development of recognition procedures for dialogue structures. The analysis of some data, like prosody are far beyond the effort for an industrial application.

To make good dialogue models affordable for many applications it is important to reduce the construction effort. Therefore, dialogue modelling tools like CSLUrp [1] neglect the variety of phenomena in spoken language and build up dialogue from a set of restricted slot-filling dialogues. Others apply a supervised learning algorithm using some dialogue structuring theory with a given set of dialogue acts [2-6].

In contrast to other learning approaches to dialog modelling DIA-MOLE does not employ theory-based dialogue units because they are subject to human interpretation and often cannot be recognized from data available in a spoken-language system. A similar approach adopting unsupervised learning for dialogue acts relies on human-labeled tags [7]. We pursue a data-driven approach and apply unsupervised learning to a sample set of spontaneous dialogues using multiple knowledge sources, i.e. domain and task knowledge, word recognition, syntax, semantics and prosodic information.

Given these data, DIA-MOLE supports segmentation of turns and interpretation of their illocutionary force based on a model of the task. As a result of learning we obtain domain- and task-specific dialogue acts (DDA) with associated features. Validations of the set of learned DDAs has shown that they are prominent for this domain and task. Dialogue act prediction was employed to evaluate our approach.

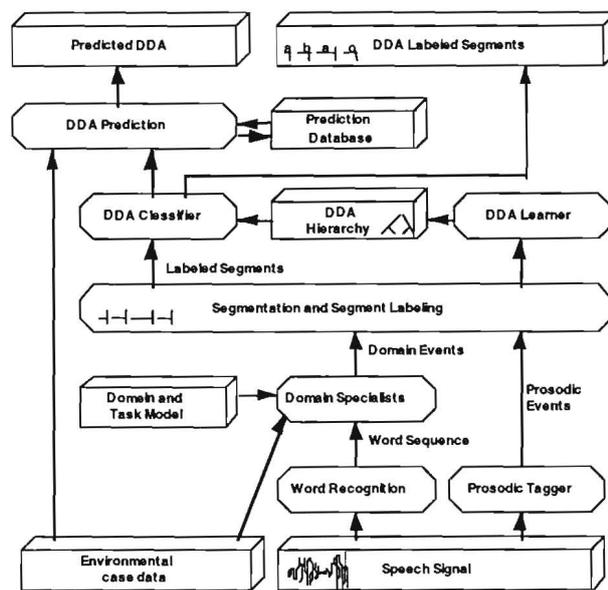


Figure 1: Architecture of DIA-MOLE

Predictions may be used by again other modules of the spoken dialogue system to adapt their environment. E.g. the word recognizer may use dialogue act predictions to choose a specific language model trained on these DDA classes to improve word recognition. Furthermore, a dialogue-planner module can use predicted DDA with their associated features as prototypes for generation to display a very natural behaviour in dialogue. A dialogue planner based on DDA predictions and case data is actually under development.

The architecture of DIA-MOLE also allows self-adapting dialogue models. If DIA-MOLE is integrated in a spoken dialogue system which is in practical use, all occurring dialogue turns, or more precisely, segments may be presented to both, the DDA classifier for further processing of that turn and to the DDA learner to improve the model in its application situation. For this reason we adopted an incremental learning algorithm within the DDA learning module.

3 System overview

Word recognizer [8] and prosodic tagger [9] are modules that were developed independent of this work. Domain and task model allow us to interpret the intention of the communicative agent's utterances. For the domain of appointment scheduling we first did an analysis of the problem and a typical problem solving method using the modeling method KADS [10, 11] (see Figure 2).

Domain specialists carry out syntactic and semantic analysis according to the domain structure and the underlying task model. Spoken language does usually not consist of well-formed sentences [12, 13], thus syntactic parsing in DIA-MOLE is restricted to partial parsing. As a representation for lexical semantics we use Conceptual Graphs [14]. Entries to lexical semantics are automatically derived from the WordNet ontology. Domain specialists provide as a result domain-relevant events.

Segmentation breaks turns into segments which are labeled with features from prosody (accent, phrase boundaries, sentence modality and focus [15]) and domain events.

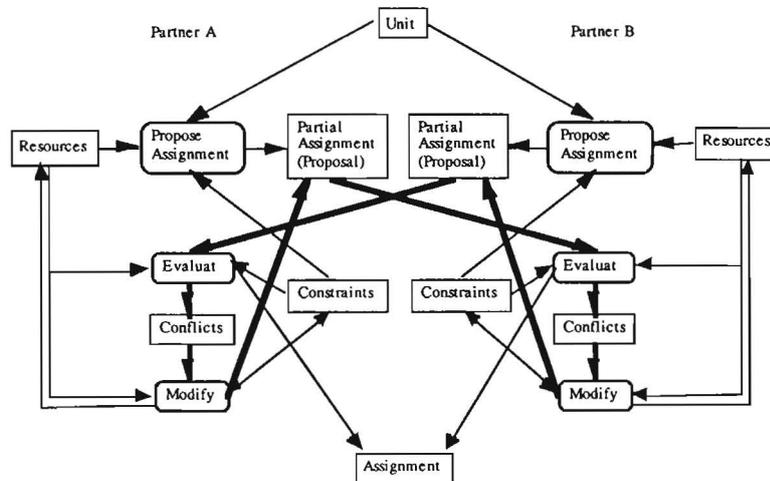


Figure 2: Functional structure of interactive appointment scheduling

These labeled segments (see for example Figure 3) are presented to the learning algorithm CLASSITALL resulting in a classification hierarchy of DDAs.

While the authors of most learning algorithm concentrate on how to learn classification hierarchies, they do not particularly address how to use their classification hierarchy. When classifying cases in DIA-MOLE, we apply three pruning methods to the resulting classification path: (a) as an absolute boundary the minimal number of cases that were classified into that class, (b) as a relative boundary the maximal case prediction and (c) a good prediction gain value from one level of the classification hierarchy to the next level.

The resulting DDA hierarchy can be used to classify and automatically label segments. From automatically labeled dialogues, a prediction module learns about predicting a subsequent act in a dialogue. Predictions are based on a ngram-model ($n \leq 3$). DIA-MOLE actually runs in SICStus- and Quintus-Prolog on different platforms.

4 CLASSITALL - an integrated learning algorithm for different kind of data

For the classification of utterance segments into pragmatical oriented dialogue act classes we use CLASSITALL [16]. It is an incremental, polythetic and unsupervised learning algorithm based on COBWEB [17] or its decendent CLASSIT [18], respectively. CLASSITALL integrates numeric and symbolic values and adds features for dealing with uncertain and incomplete knowledge. The latest version is even able to process Conceptual Graphs as structured values within the same framework.

CLASSITALL incrementally builds up a classification hierarchy from a set of cases. A case description is a data set consisting of a set of attribute-value pairs. Nodes or classes in the classification hierarchy will be called categories. For abstracting categories, COBWEB [17] uses a measure of clustering quality. From this quality measure, a

```

attitude (POSITIVE | NEGATIVE)
location (LOCAL | GLOBAL)
conflict (CONFLICT)
date-and-time-interval (SAME | NEW | ALTERNATIVE)
date-and-time-specificity (SPECIFY | GENERALIZE | SAME)
assignment (ASSIGNMENT)
phonMod (QUERY | ASSERTION | CONTINUATION)
turn (EXIT)

```

Figure 3: List of segment features that are possible in this implementation

value called *category utility* is derived by transformation and normalization that allows to determine where to integrate new cases. CLASSIT [18] replaced discrete values from COBWEB with numeric values using about the same algorithm to compute the hierarchy. This paper relates on those previous works without reciting formulas and algorithms.

4.1 Clustering of symbolic, numeric, uncomplete and uncertain data

In CLASSITALL we adopt the attribute-value paradigm from COBWEB resp. CLASSIT and integrated numeric and symbolic values. For this purpose, we had to normalize numeric values. The category value for numeric values is then modified to

$$\sum_k P(C_k) \cdot \left(\sum_i \frac{1}{\sigma_{ik} 2\sqrt{\Pi}} - \sum_i \frac{1}{\sigma_{ip} 2\sqrt{\Pi}} \right) / K$$

Function 1: Normalized category utility for numeric values in CLASSITALL¹

The expressiveness of the data that could be processed by the learning algorithm was extended towards the needs of dialogue modelling and spoken language. Input of probability values for attribute-value pairs enables us to use probability values stemming from underlying modules in a spoken language system directly for processing within the learning algorithm. They may express the quality of data. Moreover, it is possible to exert influence on the importance of specific attributes and values for the resulting classification hierarchy.

$$P(A_i = V_{ij}) \Leftrightarrow \frac{\sum_c (A_{ci} = V_{ij}) Q_{ij}}{\sum_c A_{ci} Q_{ix}}, \text{ where } Q_{ab} = A Q_{ab} C Q_{ab}$$

Function 2: Attribute-value probability for symbolic values in CLASSITALL

Computation of the symbolic and numeric category utility had to be changed to incorporate these extensions. In Function 2 AQ is the quality of an attribute having one

¹ σ_{ik} is the standard deviation of an attribute in a given category, σ_{ip} is the standard deviation for an attribute in its parent category, and K is the number of cases

specific value and CQ the quality of a whole case, both given as probability values, c stands for cases. Similar changes had to be made for numeric values. For numeric values mean and standard deviation in CLASSITALL are weighted with the quality Q_i resulting in Functions 3, where $Q_i=AQ_i Cq_i$.

$$\mu\omega = \frac{\sum_i V_i Q_i}{\sum_i Q_i} \quad \text{and} \quad \sigma\omega = \sqrt{\frac{\sum_i (\mu\omega - V_i Q_i)^2}{\sum_i Q_i}}$$

Functions 3: Weighted mean and weighted standard deviation for numeric attributes in CLASSITALL

4.2 Clustering structured values

CLASSITALL also integrates structured values, represented in the Conceptual Graphs formalism [14]. The development of an efficient algorithm for classifying structured values according to the Conceptual Graph formalism enables us to integrate syntactic and semantic structures into the learning of DDAs.

Classification of semantic networks typically means to build up an overall subsumption hierarchy like a taxonomy for KL-One terms [19] or a hierarchy of Conceptual Graphs [20, 21]. In contrast to other approaches to conceptual clustering of structural knowledge [22-24] that build up a unique classification tree, thus applying monothetic learning, we stick to the COBWEB paradigm of polythetic clustering and allow for multiple complex attribute clustering.

Instead of classifying a graph into a complete graph hierarchy CLASSITALL a) provides with every category a list of set-up points within a virtually existing, but not computed overall graph hierarchy that should be tested, and b) construct a partial generalisation space consisting only of maximally extended common generalisations.

For integration of structured values into CLASSITALL one major problem is to compare them. While symbolic values are easy to compare - either they are equal or not - the similarity of numeric values is harder to determine. Similarity for real-numeric values could be computed by discretizing them into some intervals (c.f. [25]). Gennari et al. [18] chose another approach and stored a continuous normal distribution for each attribute, expressed and computed using mean value μ and standard derivation σ , causing some changes to the original COBWEB algorithm to deal with the numeric category utility, i.e. *cutoff* and *acuity*. Comparison of structured values is even more

case	attribute	value
A	domain	[Appointment]->(Temporal-Locating)->[Time-Interval]->(Start-Time)->[Time-Point].
B	domain	[Appointment]->(Temporal-Locating)->[Time-Interval]->(End-Time)->[Time-Point].
E	domain	[Appointment]->(Spacial-Locating)->[Location].

Figure 4: Categories describing a single case

complicated, because structured values may consist of arbitrarily overlapping graphs. Usually, the degree of overlapping of graphs is regarded as a measure of semantic similarity between graphs. The category utility for clustering of graphs in a hierarchy should be dependent on how much graphs in a category share with each other.

Let us have a look at an example from the domain of appointment scheduling and reduce the semantics of utterances of dialogue partners to a simple description of domain-relevant parts. If we want to cluster these utterances using only the *domain* attribute (Figure 4), we would expect to be case A and B in the same category, while E should be in an other one. The algorithm and internal representation has been described elsewhere [16].

5 Combining and integrating different learning methods for NLP

In DIA-MOLE different methods for learning and natural language processing are used on different levels of processing.

5.1 Statistics for recognition from the speech signal

For determining prosodic events a statistical component using a Gaussian distribution classifier over a set of 30 energy features from the speech signal is applied. This is further described in [9].

Hidden Markov models are used to describe automata for word recognition. The HTK toolset is used to learn and apply these models [8]. Possible word graphs are weighted and pruned by knowledge on probable word sequences given by n-grams.

Both approaches do not return simple symbolic results, but a set of solutions weighted by probability. It seems to be reasonable to consider the quality of the results indicated by their probability with processing on higher level. For a parser in a dialogue system specialised on analyzing relevant parts of an utterance it has turned out that it offers the same performance when it is based on either the best chain or a whole word graph.

5.2 Syntactic and meaning analysis

A pure symbolic approach is applied in those cases, where grammar rules are known beforehand, or where distinct actions have to be performed. In the domain of appointment scheduling this is e.g. the parsing and interpretation of time expressions and the intertwined update of the dialogue context. The advantage of this manual engineered approach is that known surface constructs and classes of these can be considered even if they are not present in the available corpus.

Statistical and trainable grammar instead should be used, when there is a great variety of surface realisation that could be well trained from a corpus. This technique has not yet been used within the framework, but seems to be adequate e.g. for the recognition of more abstract constructs like *conflict* which on the other hand are involved in less interpretation.

5.3 Pragmatics and dialog processing using CLASSITALL

For the classification of utterance segments into pragmatic oriented dialogue act classes we use CLASSITALL. Different kind of data are integrated at that stage. Figure 3 shows the set of *symbolic* domain events and prosodic events that are attached as features to segments. Figure 5 gives an example for the semantics of a sentence represented using Conceptual Graphs. Concept and relation names accord to CG-PENMAN [26]. Figure 6 shows the input to CLASSITALL. In practice, semantics is not derived for whole utterances because we do not perform an overall parsing, but partial parsing. Instead some smaller graphs often just containing a single concept or a relation with two concepts are present in our data. Nevertheless, even clustering of single CG concepts is different from simple symbolic values, as a subsumption hierarchy is used while clustering.

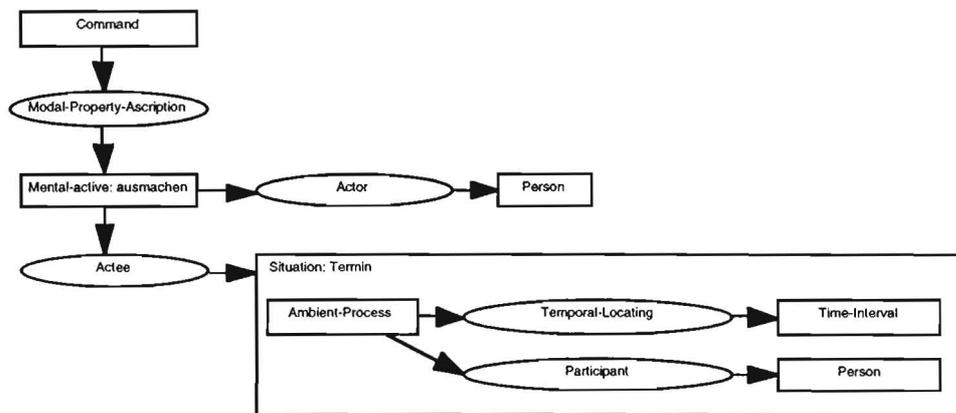


Figure 5: Semantic representation of the sentence *ja prima dann lassen Sie uns doch noch einen Termin ausmachen wann w"are es Ihnen denn recht* (Ok, let us schedule an appointment when would you like)

\$segment	1.0	n001k-0-0
attitude	symbolic	1.0 POSITIVE
assignment	symbolic	1.0 ASSIGNMENT
phonMod	symbolic	0.89 <QUESTION>
semantics	structure	1.0 [Command]-> (Modal-Property-Ascription)-> [Mental-Active: ausmachen] - ->(Actee)-> [SITUATION: Termin [Ambient-Process] - ->(Temporal-Locating) -> [Time-Interval] ->(Participant) -> [Person],, ->(Actor) -> [Person],.
text	comment	ja prima dann lassen Sie uns doch noch einen Termin ausmachen wann w"are es Ihnen denn recht

Figure 6: Feature description of a segment presented to the learning algorithm. The columns contain attribute name, type of attribute, quality and value

6 Application domain

The domain of appointment scheduling was chosen as an application domain, because speech data and a whole speech system environment for this domain were available at our department within the VERBMOBIL project. The correctness of the word recognition that is used in DIA-MOLE is about 80%.

6.1 Domain Modelling

Starting out from the problem solving method *propose-and-revise* we developed, as a first step for the application of DIA-MOLE to this domain, a functional structure of interactive appointment scheduling. Figure 2 shows the structure with a main cycle of *propose-evaluate-modify* switching between both dialogue partners. Rectangular boxes indicate data and rounded boxes stand for processes. Though they are differentiated in a problem solving model both, data and processes may be verbalised in a natural language dialogue. A unit is one appointment to schedule, and resources are the calendars of the dialogue partners.

6.2 Domain Events and Prosodic Events

As a second step domain specialists recognising contributions to one of the data or process were developed. A proposal in this domain may consist of a time interval and a location, but usually they are underspecified. Therefore, the domain specialist for date and time expressions is coupled with a specific context model yielding significantly better results. While only 25.6% of the date and time expressions without context were non-ambiguous, with the help of context the right interpretation could be found for 84.5%.

A filter and re-interpretation module was applied to the data from the prosodic tagger. Domain events and prosodic events are given to the segmentation module.

6.3 Segmentation

Rules on segmentation of turns are based on prosodic information and domain knowledge. The major segmentation rule inserts a boundary just behind an prosodic event, that follows domain events. If there are no prosodic events, this results in bigger segments comprising multiple dialogue acts. Compared to a manual segmentation based on RST theory our rules show a precision of more than 95% for the determination of segment boundaries. Connectives are the major source of errors: At least in our corpus, prosody suggests that connectives are placed at the end of a first segment, already indicating that another segment will come. This reflects human utterance planning process. Human segmentations are probably influenced by German syntax and place a connective at the beginning of the second segment. We did not write special rules to circumvent this effect, as it does not influence further processing.

6.4 Segment Labeling

In a fourth step information from prosody and domain specialist are assigned as features to segments. Figure 3 shows the set of features and values used in this application. Context information is used to pursue moves in the domain. Probability values

- Suggesting a new time interval to consider for planning
- Emphasizing an alternative time suggestion *
- Alternative time suggestion

- Acknowledgement and more specific time suggestion
- Acknowledgement and more specific time suggestion *
- Emphasizing an option for a more specific time interval (as the only one)
- Asserted more specific time interval *
- More specific time intervals

- Acknowledgement *
- Emphasized acknowledgement *
- Emphasized disagreement *

- Positive evaluated alternative time interval usually indicating a new scheduling approach
- Evaluated and emphasized alternative time interval

- Emphasized conflict
- Conflict *
- Simply mentioning a conflict

- Disagreement *
- Suggestions of locations

- Acknowledgement and assertion of a conflict
- Demand for scheduling an appointment
- A first time suggestion

- Acknowledgement

- Turns without any domain contribution

Figure 4: Learned DDA classes

stemming from the underlying processes are also added with the features yielding case descriptions for the learning algorithm.

6.5 Learned DDA-Classes

In contrast to manually labeled dialogue acts [27], learned DDA classes distinguish for example explicit and implicit rejection. This means that they do not characterize the illocutionary force by interpretation, but their illocutionary force relies on acts in the domain and its task model.

The DDA-classes in Figure 4 are based on a set of 187 spontaneously spoken dialogues from the VERBMOBIL corpus with 4521 turns. A * marks, that another segment from the same speaker will follow. It is interesting to find conversational phenomena reflected in the classes, e.g. that disagreement very seldomly stands for its own, while agreement can do this in a dialogue. The actual set does not consider general dialogue information, e.g. on greetings which will be added in a future version. Furthermore pruning within the algorithm favors frequent segment features thus suppressing clear and distinct smaller classes. This could be circumvented by weighting features.

	<i>1st Prediction</i>	<i>1st+2nd Prediction</i>
23 classes	23.70%	38.34%
abstraction to 9 classes	32.91%	54.51%

Figure 5: Hit rate for dialogue act predictions

6.6 Prediction

We also evaluated the quality of the learned DDAs by testing whether they are well suited for dialogue act prediction. For this reason we compared our approach with prediction rates on manually labeled dialogue acts reported by

Reithinger and Maier [28]. They report a hit rate for the first prediction of 29% and 45% with the first and second together when considering every turn in the data. For learning we used again 187 dialogues with 4521 turns. Prediction rates based on an unseen test set of 79 dialogues with 1495 turns are given in Figure 5. Though bothered with word recognition errors, prediction rates are comparable to human labeled dialogue acts. The hierarchical representation allows an abstraction of dialogue acts. Abstracting just one level results in better predictions than with human labeled dialogue acts. We assume that further experiments with weighting features will result in better hit rates.

6.7 Generating system utterances using DDA classes

One problem of unsupervised learning of DDAs is that their meaning is not directly accessible to humans, since the category characteristic is described as a set of feature-value-pairs which additionally result from a maximal difference to all other categories. Although DDAs are not subject to human interpretation they can directly fulfill a useful task in a dialogue system, notably in a dialogue planner. If domain constraints demand for some specific feature values that have to be communicated to the dialogue partner, e.g. *attitude negative*, the most probable predicted DDAs can be used to plan further content, e.g. *date-and-time-interval alternative* and finally utter a very natural dialogue contribution. Notice, that DDAs do not only contain information about the propositional content, but also about prosodic features which can be used for spoken language generation. This is an important contribution of DDAs since the integration and manual editing of prosodic rules into language generation still causes tremendous problems.

7 Conclusion

DIA-MOLE learns classes of dialogue acts exclusively from automatically derived data in a spoken dialogue system. Only a domain model and domain specialists have to be developed. This had two positive effects, first, we avoided the enormous effort to label dialogues, and as a consequence of this, second we are able to use very large amounts of data for learning.

It has been shown in previous work, that these classes are different from manually labeled dialogue acts, but their properties (recognisability, predictability and use in a dialogue planner) are at least equivalent. The interpretation of DDAs is somewhat artificial as they usually are just a node number with feature probabilities and their primary goal lies in the use for other system components.

The learning algorithm CLASSITALL is introduced that is able to combine different knowledge representations: Symbolic, numeric, structured and uncertain data. By using this incremental learning algorithm it is possible to self-adapt dialogue models in use. A dialogue-planner based on such self-adapting dialogue models is topic of further research.

8 References

- [1] D. G. Novick and S. Sutton, "Building on experience: Managing spoken interaction through library subdialogues", in N. LuperFoy, v.Zanten, Ed., *Proc. Twente Workshop on Language Technology, Dialogue Management in NL-Systems*, Enschede 1996, pp. 51-60.
- [2] D. J. Litman, "Classifying Cue Phrases in Text and Speech Using Machine Learning", , *Proc. Annual Meeting of the American Association for Artificial Intelligence*, Seattle 1994, pp. 806-813.
- [3] E. V. Siegel and K. R. McKeown, "Emergent Linguistic Rules from Inducing Decision Trees: Disambiguating Discourse Clue Words", , *Proc. Annual Meeting of the American Association for Artificial Intelligence*, Seattle 1994, pp. 820-826.
- [4] W. Lehnert and S. Soderland, "Corpus-Driven Knowledge Acquisition for Discourse Analysis", , *Proc. Annual Meeting of the American Association for Artificial Intelligence*, Seattle 1994, pp. 827-832.
- [5] M. Mast, R. Kompe, S. Harbeck, A. Kiessling, H. Niemann, E. Nöth, E. G. Schukat-Talamazzini, and V. Warnke, "Dialog Act Classification with the Help of Prosody", in H. T. Bunnell and W. Idsardi, Eds., *Fourth International Conference on Spoken Language Processing*, Philadelphia, PA, USA, Oct 3-6 1996, pp. 1732-1735.
- [6] S. Wermter and M. Löchel, "Learning Dialog Act Processing", *Coling96*, Copenhagen, Denmark 1996, pp. 740-745.
- [7] T. Andernach, "A Machine Learning Approach to the Classification of Dialogue Utterances", , *NeMLaP-2*, Ankara, Turkey 1996, .
- [8] K. Hübener, U. Jost, and H. Heine, "Speech recognition for spontaneously spoken German dialogs", *4th Int. Conference on Spoken Language Processing*, Philadelphia 1996.
- [9] V. Strom, "Detection of accents, phrase boundaries, and sentence modality in German", *EUROSPEECH 1995* 1995, pp. 2039-2041.
- [10] J. Breuker and B. Wielinga, "Models of Expertise", , *Proc. of the European Conference on Artificial Intelligence ECAI-86* 1986, pp. 306-318.
- [11] U. Sundin, "Assignment and Scheduling", in *CommonKADS Library for Expertise Modelling - Reusable Problem Solving Components*, J. V. d. V. Breuker, W., Ed. Amsterdam: IOS Press, 1994, pp. 231-264.
- [12] V. Weber and S. Wermter, "Towards Learning Semantics of Spontaneous Dialog Utterances in a Hybrid Framework", , *AISB95* 1995, .
- [13] W. Eckert, "Understanding of Spontaneous Utterances in Human-Machine-Dialog", in N. LuperFoy, v.Zanten, Ed., *Proc. Twente Workshop on Language Technology, Dialogue Management in NL-Systems*, Enschede 1996, pp. 139-148.
- [14] J. F. Sowa, *Conceptual Structures - Information Processing in Mind And Machine*. Reading, MA: Addison-Wesley, 1984.
- [15] A. Elsner and A. Klein, "Erkennung des prosodischen Fokus und die Anwendung im dialogaktbasierten Transfer", *Verbmobil VM-Memo 107*, 1996.
- [16] J.-U. Möller, "ClassItAll: Incremental and Unsupervised Learning in the Dia-MoLE Framework", in W. Daelemans, T. Weijters, and A. van der Bosch, Eds., *Workshop Notes of the ECML/MLnet Workshop on Empirical Learning of Natural Language Processing Tasks*, Prague, April 1997, pp. 95-104.

- [17] D. H. Fisher, "Knowledge Acquisition via Incremental Conceptual Clustering", *Machine Learning*, pp. 139-172, 1987.
- [18] J. H. Gennari, P. Langley, and D. H. Fisher, "Models of Incremental Concept Formation", *Artificial Intelligence*, pp. 11-61, 1989.
- [19] W. A. Wood, "Understanding Subsumption and Taxonomy: A Framework in Progress", in *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, J. F. Sowa, Ed.: Morgan Kaufmann, 1991, pp. 45-94.
- [20] R. Levinson, "Multi-Level Hierarchical Retrieval", in E. C. Way, Ed., *6. Annual Workshop on Conceptual Graphs*, SUNY Binghamton, 1991 1991, pp. 67-81.
- [21] G. Ellis, "Compiled Hierarchical Retrieval", in *Conceptual Structures - current research and practice*, T. E. N. Nagle, J. A.; Gerholz, L. L.; Eklund, P. W., Ed. New York, London: Ellis Horwood, 1992, pp. 271-293.
- [22] J.-U. Kietz and K. Morik, "A Polynomial Approach to the Constructive Induction of Structural Knowledge", *Machine Learning*, pp. 193-217, 1994.
- [23] I. Bournaud and J.-G. Ganascia, "Conceptual Clustering of Complex Objects: A Generalisation Space based Approach", in E. e. al., Ed., *Proc. International Conference on Conceptual Graphs*, 1995, pp. 173-187.
- [24] G. W. Mineau and R. Godin, "Automatic Structuring of Knowledge Bases by Conceptual Clustering", *IEEE Transactions on Knowledge and Data Engineering*, vol. 7, pp. 824-829, 1995.
- [25] M. Lebowitz, "Categorizing numeric information for generalisation", *Cognitive Science*, vol. 9, pp. 285-309, 1985.
- [26] R. Schirdewan, "Transformation des PENMAN Upper Models in Conceptual Graphs", Studienarbeit. University of Hamburg, 1996.
- [27] S. Jekat, A. Klein, E. Maier, I. Maleck, M. Mast, and J. J. Quantz, "Dialogue Acts in VERBMOBIL", Verbmobil VM-Report 65, 1995.
- [28] N. Reithinger, Dr. and E. Maier, "Utilizing Statistical Dialogue Act Processing in VERBMOBIL", Verbmobil VM-Report 80, 1995.

Nichtmonotones Schließen und neuronale Netze

Reinhard Blutner, Humboldt-Universität Berlin

Kurzfassung

Der Beitrag faßt die Integration von symbolischen und subsymbolischen Ansätzen als ein Problem der *Vereinheitlichung* zweier anscheinend ganz unterschiedlicher Paradigmen auf. Die Zielstellung ist also nicht, zwei unterschiedliche Verarbeitungssysteme über eine geeignete Schnittstelle miteinander zu verkoppeln, sondern vielmehr, eine vereinheitlichende Darstellung zu finden, bei der symbolische und subsymbolische Verarbeitung als unterschiedliche *Perspektiven* einunddesselben kognitiven Vorgangs gedeutet werden können.

Die Grundidee wird am Beispiel von Hopfield-Netzen demonstriert. Zunächst wird gezeigt, daß Aktivierungszustände als Informationszustände aufgefaßt werden können, die durch eine Präzisionsrelation strukturiert sind. Diese Relation stiftet unter gewissen Bedingungen einen deMorganschen Verband. Aktivierungszustände können demnach als Propositionen betrachtet werden. Weiter wird gezeigt, daß sich der Prozeß der Aktivierungsausbreitung (*asynchrones Updating*) als nichtmonotone Inferenz beschreiben läßt, wobei der zugrundeliegende Inferenzbegriff eine kumulative Logik begründet. Nimmt man eine *lokalistische Repräsentation* der Aktivierungszustände an (dabei wird eine transparente Zuordnung zwischen den Einheiten des Netzes und den atomaren Symbolen der Repräsentationssprache realisiert), dann läßt sich zeigen, daß es eine eindeutige Zuordnung zwischen der Gewichtsmatrix von Hopfield-Netzen und einer (mit Gewichten versehenen) Datenbasis in Poole-Systemen gibt. Bezogen auf diese Korrespondenz bilden sich das asymptotische Verhalten der Aktivierungsausbreitung und Inferenzen in (gewichteten) Poole-Systemen wechselseitig aufeinander ab.

Allgemeine Schlußfolgerungen: (i) Die demonstrierte Extraktionsmethode ermöglicht es einem Benutzer, die "Schlüsse" eines konnektionistischen Systems unmittelbar nachzuvollziehen. Sie werden als nichtmonotone Schlüsse in einer kumulativen Logik faßbar. (ii) Die Explosion des Verarbeitungsaufwands in traditionellen nichtmonotonen Systemen bei großen Datenbasen läßt sich, jedenfalls unter gewissen Bedingungen, durch den Einsatz stochastischer Prozesse ("simulated annealing") vermeiden (implementativer Konnektionismus). (iii) Die vorgeschlagene Sichtweise liefert theoretische Aufschlüsse darüber, welche Arten von Logiken zur Beschreibung emergenter Eigenschaften neuronaler Netze dienen können.

1 Einleitung

Beim Vorliegen zwei einander widersprechender Theorien gibt es mindestens vier Möglichkeiten der Konfliktbewältigung: (i) *Eine* Theorie setzt sich durch, indem sie die anderen Theorien eliminiert; (ii) *eine* Theorie setzt sich durch, indem sie gewisse Merkmale der anderen Theorie in sich aufnimmt; (iii) beide Theorien erkennen wechselseitig ihre Gültigkeit an, allerdings beschränkt auf gewisse Anwendungsdomänen, und sie versuchen gemeinsam zur Lösung komplexer Probleme beizutragen; (iv) beide Theorien werden im Rahmen einer neuen Theorie vereinheitlicht (Diese Vereinheitlichung kann in einer partiellen Reduktion bestehen: Reduktion der klassischen Thermodynamik im Rahmen der statistischen Physik. Oder sie setzt eine Neuschöpfung des theoretischen Paradigmas voraus, wobei sich die ehemals widersprechenden Theorien als unterschiedliche "Bilder" derselben Wirklichkeit erweisen: Auflösung des Welle-Teilchen-Dualismus im Rahmen der Quantenmechanik). Bezogen auf das Verhältnis zwischen Symbolverarbeitung und Konnektionismus nennen Smolensky u.a. (1992) diese vier Positionen (i) eliminativer Konnektionismus, (ii) implementativer Konnektionismus, (iii) hybrider Ansatz, (iv) integrativer Konnektionismus.

Ich möchte aus dreierlei Gründen die letztgenannte Position einnehmen. Erstens ermöglicht es diese Position, interessante Wechselbeziehungen zwischen Konnektionismus und algebraischer Semantik bzw. nichtmonotoner Logik sichtbar zu machen und für die Analyse subsymbolischer Verarbeitungssysteme zu nutzen. Zweitens erhalten gewisse Systeme der nichtmonotonen Logik dadurch eine besondere Fundierung und Auszeichnung gegenüber anderen Systemen. Und drittens eröffnet dieser Zugang die Möglichkeit, bei der Implementation (nichtmonotoner) symbolischer Inferenzsysteme auch auf Techniken subsymbolischer Systeme zurückzugreifen (z.B. stochastische Optimierungstechniken wie *simulated annealing*).

Der Plan für das weitere Vorgehen ist folgender. In Abschnitt 2 fasse ich

Aktivierungszustände in Hopfield-Netzen als Informationszustände auf und definiere eine Präzisionsrelation. Diese Relation erweist sich als eine Halbordnung, die einen deMorganschen Verband stiftet. Abschnitt 3 untersucht den Prozeß der Aktivierungsausbreitung (*asynchrones Updating*) im asymptotischen Grenzfall. Es wird gezeigt, daß sich dieser Prozeß durch einen nichtmonotonen, kumulativen Inferenzbegriff beschreiben läßt. Abschnitt 4 führt in kostenbasierte Poole-Systeme ein und beschreibt ihre Semantik mit Hilfe präferentieller Modelle. In Abschnitt 5 werden die Aktivierungszustände des Systems durch Ausdrücke einer elementaren Aussagenlogik repräsentieren. Auf dieser Grundlage wird der Zusammenhang zwischen der Kodierung von "Wissen" in der Verbindungsmatrix konnektionistischer Netze und der Repräsentation von Wissen in Poole-Systemen untersucht. Den Abschluß bildet Abschnitt 6 mit einer kurzen Betrachtung über den Wert derartiger Untersuchungen für die Verarbeitung natürlicher Sprachen.

2 Informationszustände in Hopfield-Netzen

Wir betrachten ein Hopfield-Netz mit dem Arbeitsbereich $[-1,+1]$ für die einzelnen neuronalen Einheiten (+1: maximales Feuern; 0: Ruhezustand; -1: minimales Feuern).

Zustandsraum für n Neurone: $S = [-1, 1]^n$

Verbindungsmatrix: $-1 \leq w_{ij} \leq 1$, $w_{ij} = w_{ji}$, $w_{ii} = 0$

asynchrones Updating:

$$(1) \quad s_i(t+1) = \begin{cases} \theta(\sum_j w_{ij} \cdot s_j(t)), & \text{falls } i = \text{random}(1,n) \\ s_i, & \text{sonst} \end{cases}$$

Wenn wir voraussetzen, daß sich das System nur in Richtung höherer (positiver oder negativer) Aktivierung entwickelt, dann lassen sich Aktivierungszustände als Informationszustände deuten, die entsprechend ihrem Informationsgehalt geordnet sind:

Definition 1

$\langle S, \geq \rangle$ ist ein Poset von Informationszuständen gdw.

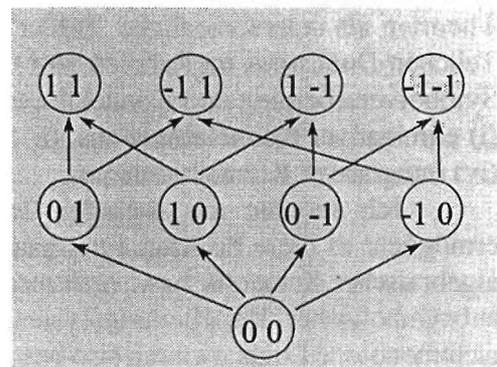
(i) $S = [-1, +1]^n$ (Zustandsraum)

(ii) $s, t \in S$: $s \geq t$ gdw. $s_i \geq t_i \geq 0$ oder $s_i \leq t_i \leq 0$, für alle $1 \leq i \leq n$.

Das Bestehen der Relation $s \geq t$ drückt aus, daß s informationell mindestens so reich (oder präzise) wie t ist. Als Element von Informationszuständen drückt der Wert 0 Unterspezifizierung aus, 1 und -1 drücken vollständige Spezifizierung aus. Informationszustände, deren Elemente nur 1 und -1 enthalten, heißen *total*.

Beispiel: Poset von Informationszuständen für $n=2$.

Dieses Poset bildet allerdings noch keinen Verband. Jedoch ist es leicht möglich, das Poset zu einem Verband zu erweitern, wenn man "uneigentliche" Aktivierungen und "uneigentliche" oder "absurde" Informationszustände einführt. Ich schreibe "nil" für uneigentliche Aktivierungen.



Definition 2

$\langle S \cup \perp, \geq \rangle$ ist ein erweitertes Poset von Informationszuständen gdw.

(i) $S = [-1, 1]^n$ (Zustandsraum)

(ii) $\perp = \{s: s_i = \text{nil für ein } 1 \leq i \leq n\}$ (Menge der *absurden* Informationszustände)

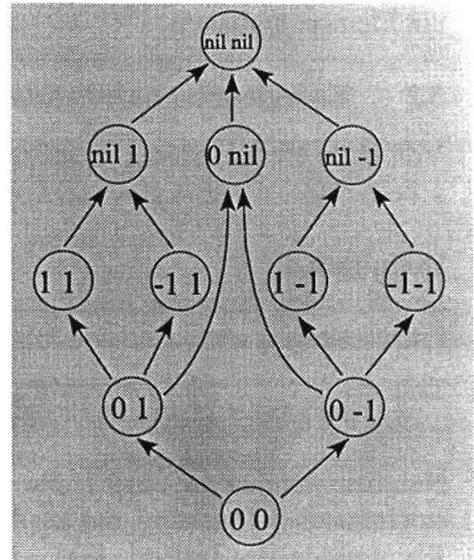
(iii) $s, t \in S \cup \perp$: $s \geq t$ gdw. $s_i = \text{nil}$ oder $s_i \geq t_i \geq 0$ oder $s_i \leq t_i \leq 0$, für alle $1 \leq i \leq n$.

Beobachtung 1

Das erweiterte Poset von Informationszuständen $\langle S \cup \perp, \geq \rangle$ bildet einen deMorganschen Verband.

Dabei kann die Konjunktion $s \circ t = \sup\{s, t\}$ als *simultane Realisierung* der Zustände s und t verstanden werden, die Disjunktion $s \circ t = \inf\{s, t\}$ als eine Art von *Generalisierung* und das Komplement s^* reflektiert das Fehlen von Information.

Anmerkung: Diese Beobachtung ist die Verallgemeinerung eines Resultats von Balkenius & Gärdenfors (1991). Die Autoren untersuchen den binären Fall und finden eine Boolesche Algebra.



3 Asymptotische Updates von Informationszuständen

Sei f eine asynchrone (stochastische) Update-Funktion (vgl. den Ausdruck (1) für den Fall diskreter Zeit). Hopfield (1982) hat gezeigt, daß dann für das im vorigen Abschnitt charakterisierte System eine Ljapunov- oder Energiefunktion existiert, und zwar

$$(2) \quad E(s) = -\sum_{i>j} w_{ij} \cdot s_i \cdot s_j.$$

Daraus ergibt sich, daß der Grenzwert $\lim_{n \rightarrow \infty} f^n(s)$ für jedes $s \in S$ existiert und einen stabilen Zustand (Resonanz) des Systems darstellt (lokales Energieminimum). Außerdem ist bekannt, daß im Falle gewisser stochastischer Update-Funktionen ("simulated annealing") der Zustand $\lim_{n \rightarrow \infty} f^n(s)$ ein *globales* Minimum der Energiefunktion darstellt (Cohen & Grossberg 1983).

3.1 Asymptotische Updates mit Klammerung

Im allgemeinen wird das *Updating* eines Informationszustands s einen Zustand $f \dots f(s)$ ergeben, der nicht die durch s vermittelte Information einschließt (d.h., dieser Zustand ist nicht notwendig eine Präzisierung von s). Für das Folgende ist es jedoch wesentlich, das *Updating* eines Informationszustands s als Präzisierung zu fassen. Nach Balkenius & Gärdenfors (1991) läßt sich dies immer dadurch erreichen, daß die bereits im Ausgangszustand spezifizierten Werte (1 und -1) "festgeklammert" werden. Die geschieht technisch in folgender Weise:

Definition 3

Sei f eine asynchrone (stochastische) Update-Funktion (für ein Hopfield-System mit der Verbindungsmatrix w):

Update-Funktionen mit Klammerung:

$$\begin{aligned} \underline{f}(s) &= f(s) \circ s; \\ \underline{f}^{n+1}(s) &= f(\underline{f}^n(s)) \circ s \end{aligned}$$

Asymptotische Update-Funktionen mit Klammerung

$$\text{ASUP}_w(s) = \{t: t = \lim_{n \rightarrow \infty} \underline{f}^n(s)\}$$

Der zufällige Charakter des *Updates* bedingt, daß mehrere asymptotische *Updates* eines Ausgangszustands s auftreten können. Die Menge $ASUP_w(s)$ wird also im allgemeinen mehr als ein Element enthalten.

3.2 Energie-minimale Präzisierungen von Informationszuständen

Von einer etwas anderen Blickrichtung können wir diejenigen Präzisierungen eines Zustands s betrachten, die eine gewisse Kostenfunktion E minimieren:

Definition 4

Sei $\langle S, \geq \rangle$ ein Poset von Informationszuständen, E eine reelle Funktion auf S . Die E -minimalen Präzisierungen von s sind wie folgt definiert:

$$\min_E(s) = \{t: t \geq s \text{ und es gibt kein } t' \geq s, \text{ so daß } E(t') < E(t)\}$$

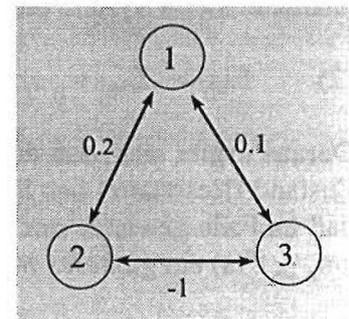
Nehmen wir nun an, daß E die Ljapunov-Funktion (2) des Systems ist und betrachten wir stochastisches *Updating*, das ein "simulated annealing" realisiert. Dann wird, auch im Falle von Klammerung, der Zustand $\lim_{n \rightarrow \infty} f^n(s)$ immer ein *globales* Minimum der Funktion E darstellen. Wir gelangen damit zu folgender

Beobachtung 2

Asymptotische *Updates* und E -minimale Präzisierungen eines Zustands s stimmen überein, d.h.: $ASUP_w(s) = \min_E(s)$

3.3 Elementares Beispiel

$$w = \begin{pmatrix} 0 & 0.2 & 0.1 \\ 0.2 & 0 & -1 \\ 0.1 & -1 & 0 \end{pmatrix}$$



Ich betrachte den Anfangszustand $\langle 1 \ 0 \ 0 \rangle$ und seine möglichen Präzisierungen:

$\{t: t \geq \langle 1 \ 0 \ 0 \rangle\}$	=	{	$\langle 1 \ 0 \ 0 \rangle$	E
			$\langle 1 \ 0 \ 1 \rangle$	0
			$\langle 1 \ 0 \ -1 \rangle$	-0.1
			$\langle 1 \ 1 \ 0 \rangle$	0.1
			$\langle 1 \ 1 \ 0 \rangle$	-0.2
			$\langle 1 \ -1 \ 0 \rangle$	0.2
			$\langle 1 \ 1 \ 1 \rangle$	0.7
Daraus ergibt sich:			$\langle 1 \ 1 \ -1 \rangle$	-1.1
$\min_E(\langle 1 \ 0 \ 0 \rangle) = \{\langle 1 \ 1 \ -1 \rangle\}$			$\langle 1 \ -1 \ 1 \rangle$	-0.9
			$\langle 1 \ -1 \ -1 \rangle$	1.3

3.4 Aktivierungsausbreitung als nichtmonotone Inferenz

Unsere Auffassung von Informationszuständen als propositionalen Objekten ermöglicht es, die Präzisierungsrelation \geq unmittelbar als einen strikten Inferenzbegriff zu deuten. Jedenfalls erfüllt die Relation \geq die tarskischen Forderungen für einen (strikten) Inferenzbegriff: Reflexivität,

Schnitt und Monotonie.

Balkenius & Gärdenfors (1991) haben verdeutlicht, daß es möglich ist, die asymptotische Aktivierungsausbreitung in neuronalen Netzen als nichtmonotone (schwache) Inferenz zu beschreiben. Sei also $\langle S_{U\perp}, \geq \rangle$ ein erweitertes Poset von Informationszuständen für ein System mit n Einheiten. Sei w die Verbindungsmatrix und E die Energiefunktion. Dann läßt sich eine inferentielle Relation \vdash zwischen Informationszuständen definieren, und zwar auf zweierlei Weise:

Definition 5 SCHWACHER INFERENZBEGRIFF (SIB)

- (i) $s \vdash_w t$ gdw. $s' \geq t$ für jedes $s' \in ASUP_w(s)$ (SIB basierend auf asympt. Updates)
- (ii) $s \vdash_E t$ gdw. $s' \geq t$ für jedes $s' \in \min_E(s)$ (SIB basierend auf E-minimalen Präzisis.)

Als unmittelbare Folgerung von Beobachtung 2 ergibt sich, daß die beiden Varianten den gleichen Inferenzbegriff definieren; es gilt also: $s \vdash_w t$ gdw. $s \vdash_E t$. Außerdem ist es nicht schwer, die folgenden Beobachtungen zu beweisen:

Beobachtungen 3

- (a) $s \geq t$, dann $s \vdash_w t$ (SUPRAKLASSISCH)
- (b) $s \vdash_w s$ (REFLEXIVITÄT)
- (c) wenn $s \vdash_w t$ und $s \circ t \vdash_w u$, dann $s \vdash_w u$ (SCHNITT)
- (d) wenn $s \vdash_w t$ und $s \vdash_w u$, dann $s \circ t \vdash_w u$ (SCHWACHE MONOTONIE)

Dabei besagt (a), daß alle strikten Schlußfolgerungen auch als nichtmonotone Schlußfolgerungen durchgehen. Die Bedingungen (b-d) stellen auf den Punkt die generellen Forderungen dar, die Gabbay, Makinson, Gärdenfors, Kraus, Lehmann, Magidor (und viele andere) als konstitutiv für nichtmonotone Inferenzsysteme ansehen (*kumulative* Logiken).

Zur Illustration des schwachen Folgerungsbegriffs betrachte ich das in Abschnitt 4.1 gegebene Beispiel (mit der dort definierten Verbindungsmatrix w).

$$\begin{array}{ll} \langle 1 \ 0 \ 0 \rangle \vdash_w \langle 0 \ 1 \ 0 \rangle & \text{wegen } \min_E[\langle 1 \ 0 \ 0 \rangle] = \{\langle 1 \ 1 \ -1 \rangle\} \\ \langle 1 \ 0 \ 0 \rangle \vdash_w \langle 0 \ 0 \ -1 \rangle & \text{"} \\ \langle 1 \ 0 \ 0 \rangle \circ \langle 0 \ 1 \ 0 \rangle \vdash_w \langle 0 \ 0 \ -1 \rangle & \text{wegen SCHWACHER MONOTONIE} \\ \langle 1 \ 0 \ 0 \rangle \circ \langle 0 \ 0 \ 1 \rangle \vdash_w \langle 0 \ -1 \ 0 \rangle & \text{wegen } \min_E[\langle 1 \ 0 \ 1 \rangle] = \{\langle 1 \ -1 \ 1 \rangle\} \end{array}$$

4 Kostenbasierte Poole-Systeme

In konnektionistischen Systemen ist Wissen kodiert in Form der Verbindungsmatrix oder der Energiefunktion. In symbolischen System bedient man sich üblicherweise (Default-) logisch basierter Datenbasen zur Wissensrepräsentation. Ein prominentes Beispiel für derartige Systeme bilden die von Poole (z.B. Poole 1988, 1994) ausgearbeiteten Systeme. Ich führe zunächst eine neue Variante derartiger Systeme ein, die ich *gewichtete Poole-Systeme* nenne. Außerdem entwickle ich eine Präferenzsemantik, die sich in Abschnitt 5 als das entscheidende Bindeglied erweisen wird zur Darstellung des Zusammenhangs zwischen diesen Systemen und Hopfield-Netzen. Hier und im folgenden betrachte ich die Sprache L_{At} der elementaren Aussagenlogik über einem Alphabet $At = \{p_1, \dots, p_N\}$ (von Elementarsymbolen).

4.1 Grundbegriffe (vgl. Poole 1988, 1994, Brewka 1991)

Definition 6

Ein Tripel $\langle At, \Delta, g \rangle$ heißt ein gewichtetes Poole-System gdw.

- (i) At ist eine nichtleere Menge (atomare Symbole)
- (ii) Δ ist eine Menge konsistenter Formeln der Sprache L_{At} (mögliche Hypothesen)

(iii) $g: \Delta \mapsto [0,1]$ (Gewichtsfunktion)

Sei nun $T = \langle At, \Delta, g \rangle$ ein gewichtetes Poole-System und sei α eine konsistente Formel (zur Repräsentation von Fakten). Dann führe ich folgende Begriffe ein:

Definition 7

(A) Ein Szenario von α in T ist eine Teilmenge Δ' von Δ , wobei $\Delta' \cup \{\alpha\}$ konsistent ist.

(B) Das Gewicht eines Szenarios Δ' ist

$$G(\Delta') = \sum_{\alpha' \in \Delta'} g(\alpha') - \sum_{\alpha' \in (\Delta - \Delta')} g(\alpha')$$

(C) Ein maximales Szenario von α in T ist ein Szenario, dessen Gewicht von keinem anderen Szenario (von α in T) übertroffen wird.

Damit sind wir vorbereitet, die folgende (skeptische) kumulative Ableitungsrelation zu definieren:

Definition 8

$\alpha \succ_T \beta$ gdw. β eine (strikte) Schlußfolgerung jedes maximalen Szenarios von α in T ist.

4.2 Ein elementares Beispiel

$$At = \{p_1, p_2, p_3\}$$

$$\Delta = \{p_1 \neg 0.2 p_2, p_1 \neg 0.1 p_3, p_2 \neg 1.0 \sim p_3\} \quad (\text{die Gewichte sind als Indizes notiert})$$

einige (relevante) Szenarios von p_1 :

	G
{}	-1.3
{ $p_1 \neg p_2$ }	-0.9
{ $p_1 \neg p_2, p_1 \neg p_3$ }	-0.7
{ $p_1 \neg p_2, p_2 \neg \sim p_3$ }	1.1 \Rightarrow
{ $p_1 \neg p_3, p_2 \neg \sim p_3$ }	0.9

Damit gilt: $p_1 \succ_T p_2, p_1 \succ_T \neg p_3$

4.3 Die Semantik gewichteter Poole-Systeme

Sei $T = \langle At, \Delta, g \rangle$ ein gewichtetes Poole-System mit $At = \{p_1, \dots, p_N\}$. Eine (totale) Interpretationsfunktion für L_{At} ist eine Funktion v von At in $\{-1, 1\}$. Für die Bewertung $[\alpha]_v$ einer Formel α von L_{At} relativ zur Interpretation v gelten die üblichen Klauseln:

$[\alpha \wedge \beta]_v = \min([\alpha]_v, [\beta]_v)$, $[\alpha \vee \beta]_v = \max([\alpha]_v, [\beta]_v)$, $[\sim \alpha]_v = -[\alpha]_v$. Für Interpretationsfunktionen v läßt sich ein Maß angeben, daß ausdrückt wie stark die gewählte Interpretation mit dem Hypothesensystem Δ in Konflikt gerät:

$$(3) \quad \mathcal{E}(v) = -\sum_{\alpha \in \Delta} g(\alpha) \cdot [\alpha]_v$$

Aus Gründen, die im nächsten Abschnitt offensichtlich werden, nenne ich dieses Maß die *Energie* der Interpretation. Sei nun α eine (konsistente) Formel von L_{At} . Die Begriffe *Modell* und *präferentes Modell* sind wie folgt definiert:

Definition 9

(A) Eine Interpretation v heißt ein *Modell* von α gdw. $[\alpha]_v = 1$.

(B) Ein *präferentes Modell* von α ist ein Modell, dessen Energie von keinem anderen Modell von α unterboten wird.

Damit sind alle Bausteine zusammengetragen, die erforderlich sind, um einen semantischen (kumulativen) Folgerungsbegriff zu definieren:

Definition 10

$\alpha \supseteq_T \beta$ gdw. jedes präferente Modell von α ein Modell von β ist.

Es läßt sich zeigen, daß dieser Folgerungsbegriff eine korrekte und vollständige Charakterisierung der in Definition 8 gefaßten kumulative Ableitungsrelation liefert:

Beobachtung 4

Für beliebige Formeln α und β von L_{At} gilt: $\alpha \supseteq_T \beta$ gdw. $\alpha \supseteq_T \beta$.

Der Beweis folgt der von Poole (1994) vermittelten Grundidee (für Einzelheiten vgl. <http://www2.rz.hu-berlin.de/asg/blutner/psylogic.ps>).

5. Integration kostenbasierter Poole-Systeme und Hopfield-Netze

Um symbolische und subsymbolische Verarbeitung als unterschiedliche *Perspektiven* einunddesselben (kognitiven) Vorgangs zu deuten, ist es zunächst erforderlich, Symbolstrukturen und Aktivierungszustände aufeinander zu beziehen. In Abschnitt 5.1 wird dies mit Hilfe von Standardtechniken der algebraischen Semantik demonstriert. Abschnitt 5.2 dehnt den Bezug auf die beiden Inferenzbegriffe aus.

5.1 Symbolische Repräsentation von Informationszuständen

Ich betrachte ein Hopfield-Netz mit n Einheiten und benutze die aussagenlogische Sprache L_{At} , um (einige/alle) Zustände des Netzes zu *repräsentieren*. Mit anderen Worten: Ich sehe die Sprache L_{At} als ein *symbolisches Mittel* an, um über die Aktivierungszustände des Systems zu sprechen. Üblichen Vorstellungen der algebraischen Semantik folgend, läßt sich diese Idee wie folgt ausdrücken: Die nicht-logischen Symbole der Sprache werden durch die Elemente der entsprechenden algebraischen Struktur (deMorganscher Verband) interpretiert—das sind in unserem Falle die Aktivierungszustände; die logischen Symbole der Sprache entsprechen gewissen Operationen in der Struktur ($\wedge \mapsto \circ, \vee \mapsto \oplus$; die innere Negation \sim entspricht der Operation $(-s)_i = -s_i$; sie konvertiert positive in negative Information und *vice versa*; die Komplementoperation $*$ —sie reflektiert das *Fehlen* von Information— hat bisher noch keine Entsprechung im logischen Vokabular von L_{At} gefunden; führe also ein (äußeres) Negationssymbols \neg ein). Die auf diese Weise konstituierten algebraischen Modelle der Sprache L_{At} nenne ich Hopfield-Modelle:

Definition 11

Sei $\langle S_{U\perp}, \geq \rangle$ das erweiterte Poset von Informationszuständen für ein System mit n Einheiten.

(i) $\langle S_{U\perp}, \geq, \uparrow \rangle$ ist ein *Hopfield-Modell* (für L_{At}) gdw. \uparrow eine Funktion ist, die jedem atomaren Symbol ein bestimmtes Element von $S_{U\perp}$ zuordnet und darüber hinaus die folgenden Bedingungen erfüllt:

$$\begin{aligned} \uparrow \alpha \wedge \beta \uparrow &= \uparrow \alpha \uparrow \circ \uparrow \beta \uparrow, & \uparrow \alpha \vee \beta \uparrow &= \uparrow \alpha \uparrow \oplus \uparrow \beta \uparrow \\ \uparrow \neg \alpha \uparrow &= \uparrow \alpha \uparrow *, & \uparrow \sim \alpha \uparrow &= -\uparrow \alpha \uparrow \end{aligned}$$

(ii) $\langle S_{U\perp}, \geq, \uparrow \rangle$ ist ein lokalistische Hopfield-Modell (für L_{At}) gdw. $\langle S_{U\perp}, \geq, \uparrow \rangle$ ein Hopfield-Modell ist und \uparrow folgende Zuordnungen realisiert:

$$\uparrow p_1 \uparrow = \langle 1 \ 0 \ \dots \ 0 \rangle, \quad \uparrow p_2 \uparrow = \langle 0 \ 1 \ \dots \ 0 \rangle, \quad \dots, \quad \uparrow p_n \uparrow = \langle 0 \ 0 \ \dots \ 1 \rangle.$$

(lokalistische Modelle realisieren eine transparente Zuordnung zwischen den Einheiten des Netzes und den atomaren Symbolen der Repräsentationssprache).

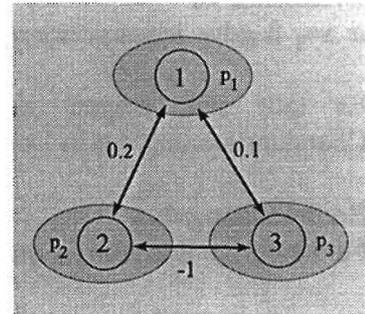
Ich will sagen, ein Informationszustand s wird durch eine Formel α vom L_{At} (relativ zu M) repräsentiert, falls $\vdash \alpha \vdash = s$ ist.

Als elementares Beispiel betrachte ich ein lokalistisches Modell für das folgende Netz:

$$\vdash p_1 \vdash = \langle 1 \ 0 \ 0 \rangle$$

$$\vdash p_2 \vdash = \langle 0 \ 1 \ 0 \rangle$$

$$\vdash p_3 \vdash = \langle 0 \ 0 \ 1 \rangle$$



Es ist unmittelbar einsichtig, daß gilt:

p_1 repräsentiert $\langle 1 \ 0 \ 0 \rangle$

p_3 repräsentiert $\langle 0 \ 0 \ 1 \rangle$

$\sim p_1$ repräsentiert $\langle -1 \ 0 \ 0 \rangle$

p_2 repräsentiert $\langle 0 \ 1 \ 0 \rangle$

$p_1 \wedge p_2$ repräsentiert $\langle 1 \ 1 \ 0 \rangle$

$p_1 \wedge p_2 \wedge \sim p_3$ repräsentiert $\langle 1 \ 1 \ -1 \rangle$

Ich nenne einen Zustand $s \in S$ *symbolisch* (relativ zu einem Hopfield-Modell M) gdw. der Zustand s durch eine Formel α in L_{At} repräsentiert wird. Es ist leicht zu sehen, daß bezogen auf ein lokalistisches Modell jeder Zustand symbolisch ist.

5.2 Übersetzung von Hopfield-Netzen in gewichtete Poole-Systeme

Für die in Abschnitt 3 untersuchten Hopfield-Netze gilt, daß für jeden (partiellen) Informationszustand u eine totale Präzisierung $t \geq u$ existiert mit $E(t) \leq E(u)$. Daraus folgt, daß die Klasse der asymptotischen *Updates* eines Aktivierungszustands s , $ASUP_w(s)$, mit jedem partiellen Informationszustand u auch eine seiner totalen Präzisierungen t enthält. Unter speziellen Bedingungen an die Netzarchitektur (keine "isolierten" Knoten etc.) läßt sich zeigen, daß $ASUP_w(s)$ nur totale Informationszustände enthält. Das besagt, daß sich unter diesen Bedingungen jeder Aktivierungszustand asymptotisch immer in total präzisierte Zustände entwickelt. (Das folgt daraus, daß für jeden (partiellen) Informationszustand eine totale Präzisierung mit *geringerer* Energie existiert)

Die Beschränkung auf Systeme, deren Dynamik asymptotisch immer zu totalen Informationszuständen führt, ermöglicht es, die nichtmonotone Inferenzrelation zwischen Aktivierungszuständen in Hopfield-Netzen (Definition 5) unmittelbar auf die Ableitungsrelation zwischen Formeln der Sprache L_{At} für gewichtete Poole-Systeme zu beziehen, jedenfalls solange eine lokalistische Betrachtungsweise von Hopfield-Netzen gewählt wird. Für den allgemeinen dynamischen Fall gelingt die Korrespondenz erst dann, wenn ein partieller Modellbegriff beim Aufbau der Poole-Systeme berücksichtigt wird. Das ist jedoch eine Aufgabe, die den gegenwärtigen Rahmen der Untersuchung sprengen würde.

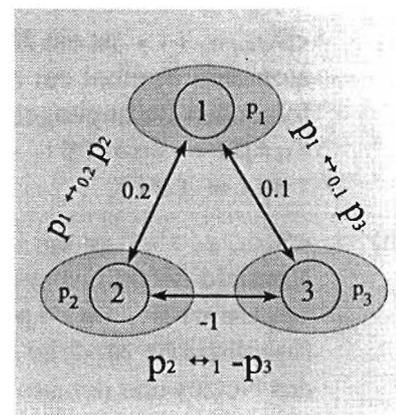
Sei nun ein Hopfield-System mit n Einheiten und der Verbindungsmatrix w gegeben. Weiter sei $At = \{p_1, \dots, p_n\}$ die Menge der atomarer Symbole der Sprache L_{At} . Ich betrachte die folgenden Formeln von L_{At} :

$$\alpha_{ij} = [p_i \leftrightarrow \text{sign}(w_{ij}) p_j], \text{ für } 1 \leq i < j \leq n$$

Für jede Verbindungsmatrix w kann damit das folgende gewichtete Poole-System $T_w = \langle At, \Delta_w, g_w \rangle$ konstruiert werden:

$$\Delta_w = \{ \alpha_{ij} : 1 \leq i < j \leq n \}$$

$$g_w(\alpha_{ij}) = |w_{ij}|$$



Es wird sich erweisen, daß diese "Übersetzung" von Verbindungsmatrizen in gewichtete Poole-Systeme zu einer strikten Korrespondenz zwischen den beiden Inferenzbegriffen führt, eine lokalistische Betrachtung von Hopfield-Netzen vorausgesetzt.

Zunächst ist es erforderlich, auf die wechselseitige Entsprechung zwischen totalen Informationszuständen und totalen Interpretationsfunktionen (vgl. Abschnitt 4.3) hinzuweisen. Jedem totalen Informationszustand t entspricht nämlich genau eine totale Interpretationsfunktion v/t mit $v/t(p_i) = t_i$ (und umgekehrt). Die folgenden Beobachtungen sind dann leicht zu bestätigen:

Beobachtungen 5

- (a) $[p_i]_{v/t} = t_i$
- (b) $[\sim \alpha]_{v/t} = -[\alpha]_{v/t}$
- (c) $[\alpha \wedge \beta]_{v/t} = [\alpha]_{v/t} \cdot [\beta]_{v/t}$
- (d) $t \geq \alpha$ gdw. $[\alpha]_{v/t} = 1$, falls die Formel α eine Konjunktion von Literalen (Atome oder deren innere Negation) ist
- (e) $\mathcal{E}(v/t) = E(t)$ (d.h. $\sum_{\alpha \in \Delta} g(\alpha) \cdot [\alpha]_{v/t} = \sum_{i > j} w_{ij} \cdot t_i \cdot t_j$)

Dabei ist E die Energiefunktion eines Hopfield-Netzes mit der Verbindungsmatrix w und \mathcal{E} ist die Energiefunktion des gewichteten Poole-Systems T_w .

Wenn die Formeln α und β Konjunktionen von Literalen sind, dann ergibt aus den Beobachtungen 5 sich daraus und aus den Definitionen 5 und 10 das folgende

Theorem

$\vdash \alpha \vdash_w \vdash \beta \vdash$ gdw. $\alpha \supseteq_T \beta$ (gdw. $\alpha \supseteq_{\neg T} \beta$).

Damit ist der Zusammenhang zwischen der (asymptotischen) Aktivierungsausbreitung in Hopfield-Netzen einerseits und nichtmonotonen Schlüssen in gewichteten Poole-Systemen andererseits ausgedrückt. Dieser Zusammenhang ist aus zweierlei allgemeinen Gründen nützlich: (i) Dank diesem Zusammenhang kann eine Schwäche konnektionistischer Systeme überwunden werden, die in der weitgehenden "Undurchschaubarkeit" der Verarbeitungseffekte derartiger Systeme besteht. Die demonstrierte Übersetzungsmethode ermöglicht es einem Benutzer, die "Schlüsse" eines konnektionistischen Systems unmittelbar nachzuvollziehen. Sie werden als symbolische, nichtmonotone Schlüsse in einer kumulativen Logik faßbar. (ii) Die Explosion des Verarbeitungsaufwands in traditionellen nichtmonotonen Systemen bei großen Datenbasen läßt sich möglicherweise durch den Einsatz stochastischer Prozesse ("simulated annealing") vermeiden (implementativer Konnektionismus).

Darüber hinaus liefert die vorgeschlagene Sichtweise theoretische Aufschlüsse darüber, welche Arten von Logiken zur Beschreibung emergenter Eigenschaften neuronaler Netze dienen können und gibt derartigen Logiken eine besondere *theoretische Rechtfertigung*.

6. Potentielle Anwendungen für die Verarbeitung natürlicher Sprachen

Geht es um die Belange der natürlichen Sprachverarbeitung, dann scheint eine unmittelbare Anwendung der hier dargestellten Überlegungen zur Integration symbolischer und konnektionistischer Ansätze auf den ersten Blick nahezu völlig ausgeschlossen. Dazu sind die gegenwärtigen Beschränkungen zu drastisch und die Vereinfachungen zu grob. Das betrifft insbesondere die vollständige Ausklammerung der Bindungsproblematik, die Beschränkung auf eine lokalistische Betrachtung von Hopfield-Netzen und die weitgehende Ignoranz gegenüber der asymptotischen Einstellung von *partiellen* Informationszuständen. Wenn die vorliegende Studie von einem gewissen Wert ist, dann besteht dieser offenbar in erster Linie darin, eine Methode zu exemplifizieren, die darauf abzielt, eine komplexe Wirklichkeit durch die

Betrachtung aus unterschiedlichen Perspektiven besser zu verstehen. Die Herstellung eines systematischen Bezugs zwischen den unterschiedlichen Perspektiven bildet den Kern eines derartigen Forschungsprogramms und unterscheidet es beispielsweise von einem hybriden Ansatz (z.B. Wermter & Lehnert 1989; für eine markante Kritik vgl. Cottrell 1989). In unseren Falle führt die integrative Methode auf interessante Wechselbeziehungen zwischen konnektionistischer Theorienbildung und einigen neueren Entwicklungen innerhalb der modelltheoretischen Semantik (Präferenzsemantik, Defaultsemantik, Datensemantik, *Update*-Semantik etc.).

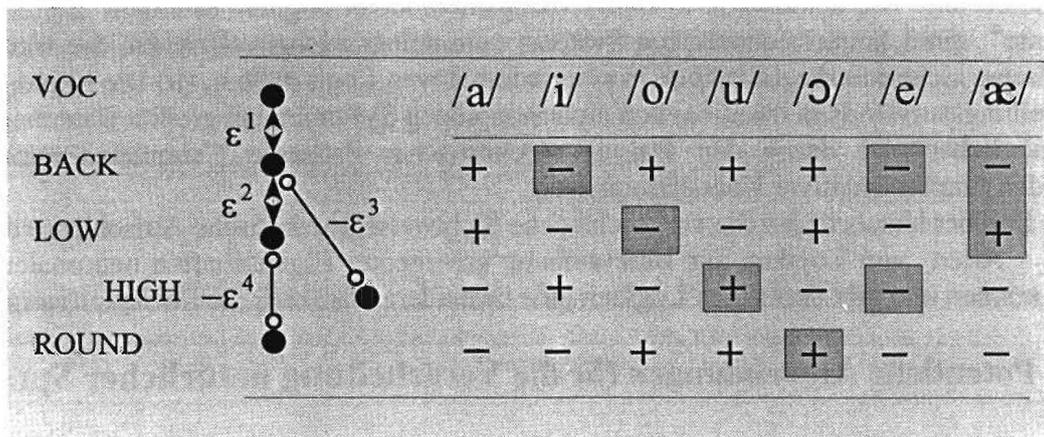
Dennoch erscheint es angemessen, auf einen Aspekt der *Semantik* und *Pragmatik* natürlicher Sprachen etwas genauer einzugehen, der in jüngster Zeit eine erhebliche Rolle in der natürlichen Sprachverarbeitung spielt. Ich meine den Aspekt der (semantischen) Unterspezifizierung und die Realisierung von Mechanismen zur "plausiblen" Vervollständigung unterspezifizierter semantischer Repräsentationen (vgl. van Deemter & Peters 1996; van der Sandt, Blutner & Bierwisch 1997).

Es ist zweckmäßig, die Idee der Unterspezifizierung, der wissenschaftsgeschichtlichen Entwicklung folgend, zunächst kurz an einem elementaren Beispiel aus der (intra-segmentalen)

-back	+back	
/i/	/u/	+high
/e/	/o/	-high/-low
/æ/	/ɔ/	+low
	/a/	

Phonologie zu erläutern. Ich betrachte das nebenstehende Segmentsystem und will eine möglichst ökonomische Beschreibung dieses Fragments durch die Merkmale BACK, LOW, HIGH, ROUND geben. Diese Merkmale können als die Atome einer elementaren Sprache L_{At} aufgefaßt werden (wobei folgende strikten Restriktionen vorausgesetzt werden: LOW \rightarrow \sim HIGH; ROUND \rightarrow BACK).

Das generische Wissen eines phonologischen Agenten hinsichtlich dieses Fragments kann durch ein Hopfield-Netz repräsentiert werden. Dabei werden exponentielle Gewichte mit der Basis $0 < \epsilon \leq 0.5$ benutzt werden (eine Annahme, die auch die Optimalitätstheorie voraussetzt). Die folgende Grafik zeigt, daß nur ein geringer Teil der Merkmalsausprägungen explizit gegeben sein muß (im Bild dunkel markiert), um eine vollständige Spezifizierung zu erreichen.



Der gleiche Effekte läßt sich durch die Zuordnung des folgenden Poole-Systems erreichen:

VOC $\rightarrow \epsilon^1$ BACK; BACK $\rightarrow \epsilon^2$ LOW
 LOW $\rightarrow \epsilon^4$ \sim ROUND; BACK $\rightarrow \epsilon^3$ \sim HIGH

Man beachte, daß diese Defaults unmittelbar als Ausdrücke der Markiertheitstheorie von Keane (1975) verstanden werden können. Die Auffüllung unterspezifizierter Merkmalsmatrizen kann

man so aus zwei Perspektiven sehen: der gewohnten markiertheits-theoretischen Perspektive und einer netzlinguistischen Perspektive. Beide Perspektiven haben offensichtlich ihre Vorzüge.

Motiviert durch das *Puzzle der kombinatorischen Explosion* haben gegenwärtige Untersuchungen zur semantischen und konzeptuellen Interpretation die Idee der Unterspezifizierung aufgegriffen mit dem Ziel, möglichst alle "nichtmonotonen" Operationen zu eliminieren, insofern diese zu Informationsverlusten bzw. zur Zerstörung einmal erzeugter semantischer Repräsentationen führen (z.B. van Deemter & Peters 1996; van der Sandt, Blutner & Bierwisch 1997). Die dabei benutzten symbolischen Techniken zur "plausiblen" Vervollständigung unterspezifizierter semantischer Repräsentationen basieren im wesentlichen entweder auf Abduktion (z.B. Hobbs et al. 1993; Blutner, Leßmöllmann, & van der Sandt 1995) oder auf bestimmten nichtmonotonen Logiken (z.B. Lascarides & Asher 1993 und verschiedene Papiere in van Deemter & Peters 1995). Es liegt in der Natur dieser symbolischen Mechanismen, daß sie wegen ihrer eigenen Komplexitätsproblematik den angestrebten Effekt teilweise oder vollständig wieder beseitigen. Die Einbeziehung konnektionistisch motivierter Verarbeitungsmechanismen (auch im Sinne des implementativen Konnektionismus, z.B. Derthick 1990) kann hier nützlich sein. Für die Klärung der dabei zahlreich auftretenden theoretischen Fragen erscheint ein integrativer Ansatz besonders vielversprechend.

Literatur

- Balkenius, C. & Gärdenfors, P. (1991): "Nonmonotonic inferences in neural networks". In J.A. Allen, R. Fikes, & E. Sandewall (eds), *Principles of knowledge representation and reasoning*. San Mateo, CA: Morgan Kaufmann.
- Blutner, R., Leßmöllmann, A., & van der Sandt, R. (1996): "Conversational implicature and lexical pragmatics". In: *Proceedings of the AAAI Spring Symposium on Conversational Implicature*. Stanford, 1-9.
- Brewka, G. (1991): *Nonmonotonic reasoning: Logical foundations of commonsense*. Cambridge: Cambridge University Press.
- Cohen, M.A. & Grossberg, S. (1983): "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks". *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, 815-826.
- Cottrell, G.W. (1989): "Research Note: A hybrid model of the intentional behavior of the dog". *Connection Science*, 3, 341-342.
- Derthick, M. (1990): "Mundane reasoning by setting on a plausible model". *Artificial Intelligence*, 46, 107-157.
- Hobbs, J.R., Stickel, M.E., Appelt, D.E., & Martin, P. (1993): "Interpretation as abduction". *Artificial Intelligence*, 63, 69-142.
- Hopfield, J.J. (1982): "Neural networks and physical systems with emergent collective computational abilities". *Proceedings of the National Academy of Sciences* 79, 2554-2558.
- Kean, M.L. (1975): *The theory of markedness in generative grammar*. Ph.D. thesis, MIT, Cambridge, Mass.
- Lascarides, A. & Asher, N. (1993): "Temporal interpretation, discourse relation, and common sense entailment". *Linguistics and Philosophy*, 16, 437-494.
- Poole, D. (1988): "A logical framework for default reasoning". *Artificial Intelligence*, 36, 27-47.
- Poole, D. (1996): "Who chooses the assumptions?". In P. O'Rorke (Ed.), *Abductive Reasoning*. Cambridge: MIT Press.
- Shoham Y. (1986): *Reasoning about change: Time and causation from the standpoint of artificial intelligence*. Ph. D. Thesis, Yale University.
- Smolensky, P. Legendre, G. , & Miyata, Y (1992): *Principles for an integrated connectionist/symbolic theory of higher order cognition*. Unpublished Paper, University of Colorado.
- van Deemter, K. & Peters, S. (1996): *Semantic Ambiguity and Underspecification*. Stanford, California: CSLI Publications.
- van der Sandt, R., Blutner, R. & Bierwisch, M. (1997): *From underspecification to interpretation*. IBM Working Papers.
- Wermter, S. & Lehnert, W.G. (1989): "A hybrid symbolic/connectionist model for noun phrase understanding". *Connection Science*, 3, 255-272.



Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

-Bibliothek, Information
und Dokumentation (BID)-
PF 2080
67608 Kaiserslautern
FRG

Telefon (0631) 205-3506
Telefax (0631) 205-3210
e-mail
dfkibib@dfki.uni-kl.de
WWW
http://www.dfki.uni-
sb.de/dfkibib

Veröffentlichungen des DFKI

Die folgenden DFKI Veröffentlichungen sowie die aktuelle Liste von allen bisher erschienenen Publikationen können von der oben angegebenen Adresse oder (so sie als per ftp erhaeltlich angemerkt sind) per anonymous ftp von ftp.dfki.uni-kl.de (131.246.241.100) im Verzeichnis pub/Publications bezogen werden. Die Berichte werden, wenn nicht anders gekennzeichnet, kostenlos abgegeben.

DFKI Publications

The following DFKI publications or the list of all published papers so far are obtainable from the above address or (if they are marked as obtainable by ftp) by anonymous ftp from ftp.dfki.uni-kl.de (131.246.241.100) in the directory pub/Publications.

The reports are distributed free of charge except where otherwise noted.

DFKI Research Reports

1997

RR-97-08

Stefan Müller

Complement Extraction Lexical Rules and Argument Attraction
14 pages

RR-97-07

Stefan Müller

Yet Another Paper about Partial Verb Phrase Fronting in German
26 pages

RR-97-06

Stefan Müller

Scrambling in German - Extraction into the *Mittelfeld*
24 pages

RR-97-05

Harald Meyer auf'm Hofe

Finding Regions of Local Repair in Hierarchical Constraint Satisfaction
33 pages

RR-97-04

Serge Autexier, Dieter Hutter

Parameterized Abstractions used for Proof-Planning
13 pages

RR-97-03

Dieter Hutter

Using Rippling to Prove the Termination of Algorithms
15 pages

RR-97-02

Stephan Busemann, Thierry Declerck, Abdel Kader Diagne, Luca Dini, Judith Klein, Sven Schmeier

Natural Language Dialogue Service for Appointment Scheduling Agents
15 pages

RR-97-01

Erica Melis, Claus Sengler

Analogy in Verification of State-Based Specifications: First Results
12 pages

1996

RR-96-06

Claus Sengler

Case Studies of Non-Freely Generated Data Types
200 pages

RR-96-05

Stephan Busemann

Best-First Surface Realization
11 pages

RR-96-04

Christoph G. Jung, Klaus Fischer, Alastair Burt
Multi-Agent Planning
Using an *Abductive*
EVENT CALCULUS
114 pages

RR-96-03

Günter Neumann
Interleaving
Natural Language Parsing and Generation
Through Uniform Processing
51 pages

RR-96-02

E.André, J. Müller , T.Rist:
PPP-Persona: Ein objektorientierter Multimedia-Prä-
sentationsagent
14 Seiten

RR-96-01

Claus Sengler
Induction on Non-Freely Generated Data Types
188 pages

1995**RR-95-20**

Hans-Ulrich Krieger
Typed Feature Structures, Definite Equivalences,
Greatest Model Semantics, and Nonmonotonicity
27 pages

RR-95-19

*Abdel Kader Diagne, Walter Kasper, Hans-Ulrich Krie-
ger*
Distributed Parsing With HPSG Grammar
20 pages

RR-95-18

Hans-Ulrich Krieger, Ulrich Schäfer
Efficient Parameterizable Type Expansion for Typed
Feature Formalisms
19 pages

RR-95-17

Hans-Ulrich Krieger
Classification and Representation of Types in TDL
17 pages

RR-95-16

Martin Müller, Tobias Van Roy
Title not set
0 pages

Note: The author(s) were unable to deliver this docu-
ment for printing before the end of the year. It
will be printed next year.

RR-95-15

Joachim Niehren, Tobias Van Roy
Title not set
0 pages

Note: The author(s) were unable to deliver this docu-
ment for printing before the end of the year. It
will be printed next year.

RR-95-14

Joachim Niehren
Functional Computation as Concurrent Computation
50 pages

RR-95-13

Werner Stephan, Susanne Biundo
Deduction-based Refinement Planning
14 pages

RR-95-12

Walter Hower, Winfried H. Graf
Research in Constraint-Based Layout, Visualization,
CAD, and Related Topics: A Bibliographical Survey
33 pages

RR-95-11

Anne Kilger, Wolfgang Finkler
Incremental Generation for Real-Time Applications
47 pages

RR-95-10

Gert Smolka
The Oz Programming Model
23 pages

RR-95-09

M. Buchheit, F. M. Donini, W. Nutt, A. Schaerf
A Refined Architecture for Terminological Systems:
Terminology = Schema + Views
71 pages

RR-95-08

Michael Mehl, Ralf Scheidhauer, Christian Schulte
An Abstract Machine for Oz
23 pages

RR-95-07

*Francesco M. Donini, Maurizio Lenzerini, Daniele Nar-
di, Werner Nutt*
The Complexity of Concept Languages
57 pages

RR-95-06

Bernd Kiefer, Thomas Fettig
FEGRAMED
An interactive Graphics Editor for Feature Structures
37 pages

RR-95-05

Rolf Backofen, James Rogers, K. Vijay-Shanker
A First-Order Axiomatization of the Theory of Finite
Trees
35 pages

RR-95-04

M. Buchheit, H.-J. Bürckert, B. Hollunder, A. Laux, W. Nutt, M. Wójcik
Task Acquisition with a Description Logic Reasoner
17 pages

RR-95-03

Stephan Baumann, Michael Malburg, Hans-Guenther Hein, Rainer Hoch, Thomas Kieninger, Norbert Kuhn
Document Analysis at DFKI
Part 2: Information Extraction
40 pages

RR-95-02

Majdi Ben Hadj Ali, Frank Fein, Frank Hoenes, Thorsten Jaeger, Achim Weigel
Document Analysis at DFKI
Part 1: Image Analysis and Text Recognition
69 pages

RR-95-01

Klaus Fischer, Jörg P. Müller, Markus Pischel
Cooperative Transportation Scheduling
an application Domain for DAI
31 pages

1994**RR-94-39**

Hans-Ulrich Krieger
Typed Feature Formalisms as a Common Basis for Linguistic Specification.
21 pages

RR-94-38

Hans Uszkoreit, Rolf Backofen, Stephan Busemann, Abdel Kader Diagne, Elizabeth A. Hinkelman, Walter Kasper, Bernd Kiefer, Hans-Ulrich Krieger, Klaus Netter, Günter Neumann, Stephan Oepen, Stephen P. Spackman.
DISCO—An HPSG-based NLP System and its Application for Appointment Scheduling.
13 pages

RR-94-37

Hans-Ulrich Krieger, Ulrich Schäfer
TDL - A Type Description Language for HPSG, Part 1: Overview.
54 pages

RR-94-36

Manfred Meyer
Issues in Concurrent Knowledge Engineering. Knowledge Base and Knowledge Share Evolution.
17 pages

RR-94-35

Rolf Backofen
A Complete Axiomatization of a Theory with Feature and Arity Constraints
49 pages

RR-94-34

Stephan Busemann, Stephan Oepen, Elizabeth A. Hinkelman, Günter Neumann, Hans Uszkoreit
COSMA - Multi-Participant NL Interaction for Appointment Scheduling
80 pages

RR-94-33

Franz Baader, Armin Laux
Terminological Logics with Modal Operators
29 pages

RR-94-31

Otto Kühn, Volker Becker, Georg Lohse, Philipp Neumann
Integrated Knowledge Utilization and Evolution for the Conservation of Corporate Know-How
17 pages

RR-94-23

Gert Smolka
The Definition of Kernel Oz
53 pages

RR-94-20

Christian Schulte, Gert Smolka, Jörg Würtz
Encapsulated Search and Constraint Programming in Oz
21 pages

RR-94-19

Rainer Hoch
Using IR Techniques for Text Classification in Document Analysis
16 pages

RR-94-18

Rolf Backofen, Ralf Treinen
How to Win a Game with Features
18 pages

RR-94-17

Georg Struth
Philosophical Logics—A Survey and a Bibliography
58 pages

RR-94-16

Gert Smolka
A Foundation for Higher-order Concurrent Constraint Programming
26 pages

RR-94-15*Winfried H. Graf, Stefan Neurohr*Using Graphical Style and Visibility Constraints for a Meaningful Layout in Visual Programming Interfaces
20 pages**RR-94-14***Harold Boley, Ulrich Buhmann, Christof Kremer*Towards a Sharable Knowledge Base on Recyclable Plastics
14 pages**RR-94-13***Jana Koehler*Planning from Second Principles—A Logic-based Approach
49 pages**RR-94-12***Hubert Comon, Ralf Treinen*Ordering Constraints on Trees
34 pages**RR-94-11***Knut Hinkelmann*A Consequence Finding Approach for Feature Recognition in CAPP
18 pages**RR-94-10***Knut Hinkelmann, Helge Hintze*Computing Cost Estimates for Proof Strategies
22 pages**RR-94-08***Otto Kühn, Björn Höfling*Conserving Corporate Knowledge for Crankshaft Design
17 pages**RR-94-07***Harold Boley*Finite Domains and Exclusions as First-Class Citizens
25 pages**RR-94-06***Dietmar Dengler*An Adaptive Deductive Planning System
17 pages**RR-94-05***Franz Schmalhofer, J. Stuart Aitken, Lyle E. Bourne jr.*Beyond the Knowledge Level: Descriptions of Rational Behavior for Sharing and Reuse
81 pages**RR-94-03***Gert Smolka*A Calculus for Higher-Order Concurrent Constraint Programming with Deep Guards
34 pages**RR-94-02***Elisabeth André, Thomas Rist*Von Textgeneratoren zu Intellimedia-Präsentationssystemen
22 Seiten**RR-94-01***Elisabeth André, Thomas Rist*Multimedia Presentations: The Support of Passive and Active Viewing
15 pages

DFKI Technical Memos**1998****TM-98-01***Christian Gerber*Bottleneck Analysis as a Heuristic for Self-Adaption in Multi-Agent Societies
16 pages**TM-97-02***Christian Gerber*Scalability of Multi-Agent Systems - Proposal for a Dissertation
49 pages**TM-97-01***Markus Perling*GeneTS: A Relational-Functional Genetic Algorithm for the Traveling Salesman Problem
26 pages**1997****TM-97-03***Hans-Jürgen Bürckert, Klaus Fischer, Gero Vierke*TelcTruck: A Holonic Fleet Management System
10 pages**TM-96-02***Harold Boley*Knowledge Bases in the World Wide Web: A Challenge for Logic Programming (Second, Revised Edition)
10 pages

TM-96-01

Gerd Kamp, Holger Wache

CTL — a description Logic with expressive concrete domains

19 pages

1995

TM-95-04

Klaus Schmid

Creative Problem Solving
and

Automated Discovery

— An Analysis of Psychological and AI Research —

152 pages

TM-95-03

Andreas Abecker, Harold Boley, Knut Hinkelmann, Holger Wache,

Franz Schmalhofer

An Environment for Exploring and Validating Declarative Knowledge

11 pages

TM-95-02

Michael Sintek

FLIP: Functional-plus-Logic Programming
on an Integrated Platform

106 pages

TM-95-01

Martin Buchheit, Rüdiger Klein, Werner Nutt

Constructive Problem Solving: A Model Construction
Approach towards Configuration

34 pages

1994

TM-94-05

Klaus Fischer, Jörg P. Müller, Markus Pischel

Unifying Control in a Layered Agent Architecture

27 pages

TM-94-04

Cornelia Fischer

PAntUDE – An Anti-Unification Algorithm for Expressing Refined Generalizations

22 pages

TM-94-03

Victoria Hall

Uncertainty-Valued Horn Clauses

31 pages

TM-94-02

Rainer Bleisinger, Berthold Kröll

Representation of Non-Convex Time Intervals and Propagation of Non-Convex Relations

11 pages

TM-94-01

Rainer Bleisinger, Klaus-Peter Gores

Text Skimming as a Part in Paper Document Understanding

14 pages

DFKI Documents

1998

D-98-03

Stephan Busemann, Karin Harbusch, Stefan Wermter(Hrsg.)

Hybride konnektionistische, statistische und regelbasierte Ansätze zur Verarbeitung natürlicher Sprache

Workshop auf der 21. Deutschen Jahrestagung für Künstliche Intelligenz, Freiburg, 9.–10. September 1997

75 Seiten

D-98-02

Andreas Abecker, Ansgar Bernardi, Knut Hinkelmann, Otto Kühn,

Michael Sintek

Techniques for Organizational Memory Information Systems

66 pages

D-98-01

Stephan Baumann, Jürgen Lichter, Michael Malburg, Heiko Maus,

Harald Meyer auf'm Hofe, Claudia Wenzel

Architektur für ein System zur Dokumentanalyse im Unternehmenskontext Integration von Datenbeständen, Aufbau- und Ablauforganisation

76 Seiten

1997

D-97-08

Christoph G. Jung, Klaus Fischer, Susanne Schacht

Distributed Cognitive Systems

Proceedings of the VKS'97 Workshop

50 pages

D-97-07

Harold Boley, Bernd Bachmann, Christian Blum, Christian Embacher, Andreas Lorenz, Jamel Zakraoui
 PIMaS:
 Ein objektorientiert-regelbasiertes System zur Produkt-Prozeß-Transformation
 45 Seiten

D-97-06

Tilman Becker, Stephan Busemann, Wolfgang Finkler
 DFKI Workshop on Natural Language Generation
 67 pages

D-97-05

Stephan Baumann, Majdi Ben Hadj Ali, Jürgen Lichter, Michael Malburg, Harald Meyer auf'm Hofe, Claudia Wenzel
 Anforderungen an ein System zur Dokumentanalyse im Unternehmenskontext
 — Integration von Datenbeständen, Aufbau- und Ablauforganisation
 42 Seiten

D-97-04

Claudia Wenzel, Markus Junker
 Entwurf einer Patternbeschreibungssprache für die Informationsextraktion in der Dokumentanalyse
 24 Seiten

D-97-03

Andreas Abecker, Stefan Decker, Knut Hinkelmann, Ulrich Reimer
 Proceedings of the Workshop „Knowledge-Based Systems for Knowledge Management in Enterprises“ 97
 167 pages

D-97-02

Tilman Becker, Hans-Ulrich Krieger
 Proceedings of the Fifth Meeting on Mathematics of Language (MOL5)
 168 pages

Note: This document is available for a nominal charge of 25 DM (or 15 US-\$).

D-97-01

Thomas Malik
 NetGLTool Benutzeranleitung
 40 Seiten

1996**D-96-07**

Technical Staff
 DFKI Jahresbericht 1995
 55 Seiten

Note: This document is no longer available in printed form.

D-96-06

Klaus Fischer (Ed.)
 Working Notes of the KI'96 Workshop on Agent-Oriented Programming and Distributed Systems
 63 pages

D-96-05

Martin Schaaf
 Ein Framework zur Erstellung verteilter Anwendungen
 94 pages

D-96-04

Franz Baader, Hans-Jürgen Bürckert, Andreas Günter, Werner Nutt (Hrsg.)
 Proceedings of the Workshop on Knowledge Representation and Configuration WRKP'96
 83 pages

D-96-03

Winfried Tautges
 Der DESIGN-ANALYZER - Decision Support im Designprozess
 75 Seiten

D-96-01

Klaus Fischer, Darius Schier
 Ein Multiagentenansatz zum Lösen von Fleet-Scheduling-Problemen
 72 Seiten

1995**D-95-12**

F. Baader, M. Buchheit, M. A. Jeusfeld, W. Nutt (Eds.)
 Working Notes of the KI'95 Workshop:
 KRDB-95 - Reasoning about Structured Objects: Knowledge Representation Meets Databases
 61 pages

D-95-11

Stephan Busemann, Iris Merget
 Eine Untersuchung kommerzieller Terminverwaltungssoftware im Hinblick auf die Kopplung mit natürlichsprachlichen Systemen
 32 Seiten

D-95-10

Volker Ehresmann
 Integration ressourcen-orientierter Techniken in das wissensbasierte Konfigurierungssystem TOOCON
 108 Seiten

D-95-09

Antonio Krüger
 PROXIMA: Ein System zur Generierung graphischer Abstraktionen
 120 Seiten

D-95-08

Technical Staff
DFKI Jahresbericht 1994
63 Seiten

Note: This document is no longer available in printed form.

D-95-07

Ottmar Lutz
Morphic - Plus
Ein morphologisches Analyseprogramm für die deutsche Flexionsmorphologie und Komposita-Analyse
74 Seiten

D-95-06

Markus Steffens, Ansgar Bernardi
Integriertes Produktmodell für Behälter aus Faserverbundwerkstoffen
48 Seiten

D-95-05

Georg Schneider
Eine Werkbank zur Erzeugung von 3D-Illustrationen
157 Seiten

D-95-04

Victoria Hall
Integration von Sorten als ausgezeichnete taxonomische Prädikate in eine relational-funktionale Sprache
56 Seiten

D-95-03

Christoph Endres, Lars Klein, Markus Meyer
Implementierung und Erweiterung der Sprache *ACCP*
110 Seiten

D-95-02

Andreas Butz
BETTY
Ein System zur Planung und Generierung informativer Animationssequenzen
95 Seiten

D-95-01

Susanne Biundo, Wolfgang Tank (Hrsg.)
PuK-95, Beiträge zum 9. Workshop „Planen und Konfigurieren“, Februar 1995
169 Seiten

Note: This document is available for a nominal charge of 25 DM (or 15 US-\$).

1994**D-94-15**

Stephan Oepen
German Nominal Syntax in HPSG
— On Syntactic Categories and Syntagmatic Relations
—
80 pages

D-94-14

Hans-Ulrich Krieger, Ulrich Schäfer
TDL - A Type Description Language for HPSG, Part 2: User Guide.
72 pages

D-94-12

Arthur Sehn, Serge Autexier (Hrsg.)
Proceedings des Studentenprogramms der 18. Deutschen Jahrestagung für Künstliche Intelligenz KI-94
69 Seiten

D-94-11

F. Baader, M. Buchheit, M. A. Jeusfeld, W. Nutt (Eds.)
Working Notes of the KI'94 Workshop: KRDB'94 - Reasoning about Structured Objects: Knowledge Representation Meets Databases
65 pages

Note: This document is no longer available in printed form.

D-94-10

F. Baader, M. Lenzerini, W. Nutt, P. F. Patel-Schneider (Eds.)
Working Notes of the 1994 International Workshop on Description Logics
118 pages

Note: This document is available for a nominal charge of 25 DM (or 15 US-\$).

D-94-09

Technical Staff
DFKI Wissenschaftlich-Technischer Jahresbericht 1993
145 Seiten

D-94-08

Harald Feibel
IGLOO 1.0 - Eine grafikunterstützte Beweisentwicklungsumgebung
58 Seiten

D-94-07

Claudia Wenzel, Rainer Hoch
Eine Übersicht über Information Retrieval (IR) und NLP-Verfahren zur Klassifikation von Texten
25 Seiten

Note: This document is no longer available in printed form.

D-94-06

Ulrich Buhrmann
Erstellung einer deklarativen Wissensbasis über recyclingrelevante Materialien
117 Seiten

D-94-04

Franz Schmalhofer, Ludger van Elst
Entwicklung von Expertensystemen: Prototypen, Typenmodellierung und kooperative Wissensentwicklung
22 Seiten

D-94-03

Franz Schmalhofer

Maschinelles Lernen: Eine kognitionswissenschaftliche
Betrachtung
54 Seiten

Note: This document is no longer available in printed
form.

D-94-02

Markus Steffens

Wissenserhebung und Analyse zum Entwicklungsprozeß
eines Druckbehälters aus Faserverbundstoff
90 pages

D-94-01

Josua Boon (Ed.)

DFKI-Publications: The First Four Years
1990 - 1993
75 pages

**Hybride konnektionistische, statistische
und regelbasierte Ansätze zur Verarbeitung
natürlicher Sprache**

Workshop auf der 21. Deutschen Jahrestagung für
Künstliche Intelligenz, Freiburg, 9.–10. September 1997

Stephan Busemann, Karin Harbusch, Stefan Wermter (Hrsg.)