LDV-FORUM

Forum der Gesellschaft für Linguistische Datenverarbeitung GLDV

LDV-Forum 12.2 (1995) Forum der Gesellschaft für Linguistische Datenverarbeitung e.V.

Herausgeber
Prof. Dr. Gerhard Knorz; Gesellschaft für Linguistische Datenverarbeitung e.V. (GLDV)
Anschrift: Fachhochschule Darmstadt, Fachbereich Information und Dokumentation (IuD), Haardtring 100, D-64295 Darmstadt; Tel.: (06151) 16-8499; Fax: (06151) 16-8980; Email: knorz@fh-darmstadt.de

Redaktion
Gerhard Knorz, Ute Hauck

Wissenschaftlicher Beirat
Dr. Hans Billing, Harald Helsen, Prof. Dr. Wolfgang Hoeppner, Prof. Dr. Gerhard Knorz, Prof. Dr. Dietmar Rösner, Prof. Dr. Ulrich Schmitz, Dr. Matthias Wermke

Erscheinungsweise
Zwei Hefte im Jahr, halbjährlich zum 31. Mai und 31. Oktober

Bezugsbedingungen
Für Mitglieder der GLDV ist der Bezugspreis des LDV-Forum im Jahresbeitrag mit eingeschlossen. Jahresabonnements können zum Preis von DM 40,- (incl. Versand), Einzelexemplare zum Preis von DM 20,- (zuzügl. Versandkosten) bei der Redaktion bestellt werden.

Editorial

Vielleicht (und hoffentlich) geht es Ihnen mit dem aktuellen Heft des LDV-Forum so wie mit einem verloren geglaubten Kind: Der Ärger steigert sich je länger man auf es wartet - aber nachdem man die Hoffnung schon aufgegeben hat, überwiegt doch die Erleichterung, wenn es ganz unvermutet auftaucht. Beginnen wir also mit einer zerknirschten Entschuldigung für eine Sequenz von Engpässen bei den Arbeitskapazitäten, die Ende 1995 bei mir selbst begann und die sich dann am IAI, dort wo die Knochenarbeit zur Vorbereitung des Drucks geleistet wird, fortgesetzt hat.

Die Konsequenz aus dem aktuellen Terminproblem und der ganz allgemein ungünstigen Lage der bisherigen Redaktionstermine wird eine generelle Umstellung der Erscheinungsweise des LDV-Forum auf die Termine April und Oktober sein! Das Jahr 1996 soll im Zuge dieser Umstellung durch eine gut gefüllte Doppelnummer Jg. 1996, Nr. 1/2 abgedeckt werden, deren rechtzeitiges Erscheinen im Oktober wohl sichergestellt ist.

Wenn Sie inzwischen mißtrauisch geworden sind und sich vom Stand der Planungen und Vorbereitungen der jeweils anstehenden Ausgaben selbst ein Bild machen wollen, dann gibt es seit Februar 1996 die Gelegenheit dazu: Unter der zugegebenermaßen etwas unhandlichen URL http://www.iud.fu.darmstadlde/iudlwwwiudlpageslpublikat-!ldvforu3.htm erwarten Sie keine Hochglanz-Ankündigungen, aber ein garantiert aktueller 'Blick in die Werkstatt' in Form von kommentierten und im Werden befindlichen Inhaltsverzeichnissen.

Und wenn Sie dann schon dabei sind, sich in Sachen Computerlinguistik im Internet zu tummeln, dann gibt es eine weitere URL, die Sie unbedingt aufsuchen sollten. Unter http://www.uni.essen.de/fb3 !linse/horne.htm empfiehlt sich der Linguistik-Server Essen (LINSE) als ein bequemer Einstieg in die gesamte virtuelle Linguistik im World Wide Web. Hinweise dazu direkt im Anschluß an dieses Editorial.

Zum gegenwärtigen Zeitpunkt noch nicht im Internet angekündigt, aber für die GLDV keineswegs weniger wichtig ist die GLDV-Herbstschule '96"die vom 23. bis 27. September 1996 an der Universität Magdeburg stattfinden wird (Ankündigung Seite 37). Titel und Veranstalter versprechen ein interessantes Programm: 'Herausforderungen für die Computerlinguistik: Multilingualität, Semantik, Diskurs.' Informationen bei Prof. Dr. Dietmar Rösner!

Mit der 'Multilingualität' greift die Herbstschule ein Schlagwort auf, das seit dem 8. November 1995 Gold wert ist, genauer beziffert: 15 Mio. ECU. Mit diesem Betrag nämlich will die Europäische Kommission in den nächsten 3 Jahren (1996-1998) ein Programm zur Förderung der Sprachenvielfalt in der Europäischen Informationsgesellschaft umsetzen. Hintergrund dieses 'Multilingual Information Society Programme' (MUS) ist, daß zunehmend elektronische Tools

Verwendung finden müssen, wenn Nutzer von den weltweit in ganz unterschiedlichen Sprachen angebotenen Informationen wirklich profitieren sollen. Wenn trotz des Bedarfs der europäische Markt in der Nutzung von Fortschritten im Bereich Language Engineering zurückhinkt, dann will die EC genau hier ansetzen und mit drei Aktionslinien Bedingungen schaffen, die zu einer Expansion der Sprachwirtschaft wie Sprach-Engineering und Übersetzungsdienstleistungen führen. Die Überschriften sind (1) Förderung des Aufbaus einer Infrastruktur für Europäische Sprachenquellen (Wörterbücher, Terminologie-Datenbanken, Grammatik-Bücher, Korpora gesprochener Sprache), (2) Mobilisierung und Ausdehnung der Sprachwirtschaft und letztlich (3) Förderung der Nutzung von entwickelten Sprachen- Tools durch den öffentlichen Sektor in Europa. Informationen dazu bei EC, DG XIII/E, MLIS Office B4/008, L-2920 Luxembourg (Fax +352430134655).

Ich schlage vor, Sie blättern das Heft nun mal durch (oder hätte ich *browsen* sagen sollen?), übersehen in keinem Fall die Konvens (Seite 39), stören sich nicht daran, daß wir bisher jahrelang eine falsche ISSN abgedruckt haben (man kann sich auf nichts verlassen), wenden sich in allen Adress-Angelegenheiten nicht an die Redaktion, sondern an den GLDV-Vorstand in Bonn und Sie vertiefen sich ins Lesen. Hoffentlich. Vielleicht höre ich ja mal auch von Ihnen?

G.K.

LINSE - Ein Linguistik-Server im Internet

Das World Wide Web im Internet wird auch für Linguisten immer interessanter. Ab sofort liefert der Linguistik-Server Essen (LINSE) ein reichhaltiges kostenloses Angebot an Informationen zur Sprachwissenschaft und zu Nachbargebieten. Der FUB VI im Fachbereich 3 der Universität GH Essen bietet nicht nur die üblichen Nachrichten über Lehrveranstaltungen, Personal und laufende Aktivitäten in Essen, sondern auch eigene Publikationen, Rezensionen, ausgewählte Seminararbeiten, Bibliographien, Lernmaterial und sogar ein kleines Preisausschreiben. Direkte Verbindungen zu Navigationshilfen und zu einer Reihe anderer Internet-Adressen mit sprachwissenschaftlichem Angebot sind eingebaut, so daß der Internet-Anfänger die Essener LINSE als bequemen Einstieg in grundsätzlich die gesamte virtuelle Linguistik im World Wide Web benutzen kann. Service und Angebot werden laufend ausgebaut. Hier die Adresse: http://www.uni.essen.de/1b3I1inse/home.htm *(Ulrich Schmitz)*

Titelgestaltung
Ute Hauck, Saarbrücken

Fachbeiträge
Unaufgefordert eingesandte Fachbeiträge werden vor Veröffentlichung von mindestens zwei ReferentInnen begutachtet. Manuskripte (dreifach) sollten daher möglichst frühzeitig eingereicht werden und bei Annahme zur Veröffentlichung in jedem Fall zusätzlich auch noch auf Diskette *3W'* als ASCII oder u. TEX-Datei übermittelt werden. Formatierungshilfen *(LDVforum.sty)* werden auf Wunsch zugesandt.

Rubriken
Die namentlich gezeichneten Beiträge geben ausschließlich die Meinung der Autoren wieder.
Einreichungen
sind - wie bei Fachbeiträgen - an die Redaktion zu übermitteln.

Redaktionsschluß
Für alle Rubriken mit Ausnahme der als Fachbeiträge eingereichten Manuskripte:
für Heft 13.1+2/96: 31. Aug. 1996

Herstellung IAI,
Saarbrücken

Druck
reha GmbH,
Saarbrücken

Auflage 400
Exemplare

Anzeigen
Preisliste und Informationen: Prof. Dr. Johann Haller, Institut für Angewandte Informationsforschung (IAI), Manin-Luther-Straße 14, D-66111 Saarbrücken; Tel.: (0681) 38951-0; Fax: (0681) 38951-40; Email: hans@iai.uni-sb.de

Bankverbindung
LDV-Forum (Prof. Haller): SaarLB Saarbrücken (EU 590 500 00)
KtoNr. 20 00 21 43

G LDV -Anschrift
Prof. Dr. Winfried Lenders, Institut für Kommunikationsforschung und Phonetik (IKP), Poppelsdorfer Allee 47, D-53115 Bonn; Tel.: (0228) 735638, Fax: (0228) 735639; Email: lenders@uni-bonn.de

MODIFIKATIONEN DES TOMITA-PARSERS FÜR ID/LP UND FEATURE GRAMMARS

Jens Woch

Abstract: Die Verwendung von ID/LP-Grammatiken und komplexen Symbolen ist bei Flektionsreichen und in der Wortstellung flexiblen Sprachen wie dem Deutschen erfolgversprechender als der Rückgriff auf kontextfreie Grammatiken (cfg). Dieser Artikel beschreibt in einem ersten Schritt die dafür notwendigen Veränderungen des Tomita-Algorithmus und dessen Tabellengenerators. In einem zweiten Schritt werden wir an einfachen Beispielen zeigen, wie diese Modifikationenfunktionieren.

1 Einführung

Der Tomita-Parser verspricht, Sprachen, die nicht durch deterministisch kontextfreie Grammatiken erzeugt werden können, mit einem ähnlichen Effizienzverhalten zu parsen, wie es für Sprachen deterministisch kontextfreier Grammatiken möglich ist. Die Wesenszüge des Tomita-Algorithmus sind a) die Variation eines Tabellengenerators für LR(k)-Parser (siehe Abschnitt 5: Glossar) und b) einige geeignete Konzepte für eine effiziente Repräsentation der im Parsing-Prozeß entstehenden Stacks und Ableitungsstrukturen. Zunächst ist die Modifikation des Tabellengenerators (nämlich das Zulassen von Mehrfacheinträgen in der Parsing-Tabelle) für eine Reihe von Problemen verantwortlich: Der Parser muß nun in der Lage sein, alle denkbaren Ableitungen eines Satzes zu repräsentieren. Die Lösung besteht darin, statt einen eigenen Stack für jede Ableitung zu verwalten, sogenannte Stacks graph-structured stacks zu verwenden, sowie die Konzepte des subtree-sharings und local-ambiguity-packings zur effizienten Verwaltung der im Parsing-Prozeß anfallenden Strukturen einzusetzen. Diese Ideen werden in [4] und [5] beschrieben.

Leider muß die der Parsing-Tabelle zugrunde liegende Grammatik das cfg-Format aufweisen. Nun hat sich die cfg in der Formulierung einer flektionsreichen Sprache mit dazu womöglich hoher Flexibilität in der Wortstellung, wie die deutsche Sprache, nicht gerade bewährt. Für das Flektionsproblem bieten sich viel eher komplexe Symbole an. (Fast) freie Wortstellungen sind darüberhinaus sehr leicht mithilfe von IDILP-Grammatiken (siehe Abschnitt 5: Glossar) aufgrund ihrer strikten syntaktischen Trennung von immediate dominance relations und linear precedence statements grammatisierbar.

2 Verarbeitung von ID/LP-Grammatiken

Hier bieten sich zwei Lösungen an: Die Übersetzung von IDILP-Grammatiken in Standard-cfg bereitet keine Schwierigkeiten, die resultierende cfg mag wie üblich ihren Weg durch den Tabellengenerator finden. Allerdings ist sie weit davon entfernt, intuitiv oder problemadäquat zu sein, denn für jede mögliche LP-Expansion wird eine cf-Regel generiert - und das können schon recht viele werden. Die andere Lösung ist übersichtlicher und dazu noch einfacher:

Der Tabellengenerator ist so zu modifizieren, daß er direkt auf ID/LP-Grammatiken angesetzt werden kann. Wenn wir von einem kanonischen LR(k)-Tabellengenerator ausgehen, wie in [2] oder [4] beschrieben, so erkennen wir, daß dieser die Tabelle in zwei Schritten erstellt: Zuerst wird für jede erlaubte Regel-Ableitung eine Kante generiert, und in einem zweiten Schritt werden auf Basis dieser Kantenmenge die Tabelleneinträge erzeugt. Die Produktion der Kanten ähnelt dabei der Produktion von active charts in Chart-Parsern. Eine Kante wird notiert als $[A \rightarrow \alpha.B\beta, a]$ mit A als linke Regelseite, a als lookahead-Symbol, α der bereits bearbeitete und $B\beta$ der unbearbeitete Teil der rechten Regelseite.

```
PROCEDURE GOTO-EDGES
Input:      A set of edges I and a symbol X ∈ V ∪ T.
Output:     A set of edges J.

 J := {[A → αX.β, a] | [A → α.Xβ, a] ∈ I}
RETURN (Closure(J))
```

Die Prozedur GOTO-EDGES generiert aus allen bereits existierenden Kanten neue Kanten, in denen α rechts um das Symbol erweitert wird, um das $B\beta$ links reduziert wird (der Pointer um eine Stelle nach rechts wandert). Sodann gibt sie die Hülle aus den neuen Kanten zurück; ihre Berechnung geschieht durch Aufruf von CLOSURE. In ihr werden für alle Regeln $B \rightarrow \tau$ neue Kanten $[B \rightarrow .\delta, b]$ mit δ als eine LP-Expansion von $B \rightarrow \tau$ erzeugt. Und das ist schon die einzige Änderung, die wir am Tomita-Parser, bzw. dessen Tabellengenerator für die Verwendung von ID/LP-Grammatiken vornehmen müssen! Während $B\beta$ ohne diese Veränderung eine lineare Symbolsequenz darstellt, nämlich die rechte Seite einer herkömmlichen cf-Regel, repräsentiert es durch diese Veränderung eine Multimenge von Symbolen, deren Elemente erlaubte LP-Expansionen darstellen (die wiederum die durch LP-statements erlaubten linearen Symbolsequenzen sind).

```
PROCEDURE CLOSURE
Input:      A set of edges I

repeat until none new edges are generated:
      For all [A → α.Bβ, a] ∈ I and all δ as a LP-expansion of B → τ
      and all b ∈ FIRST(LP-expansion(β)a):
            I := I ∪ {[B → .δ, b]}
RETURN (I)
```

Die Prozedur FIRST(βa) berechnet eine Menge von erreichbaren Terminalen.

Hier der modifizierte kanonische LR(k)-Tabellengenerator:

ALGORITHM modCLR
Input: A grammar G in ID/LP, and an additional rule $S' \rightarrow S$
Output: A parsing table for G

Calculation of $C = \{I_0, I_1, ..., I_n\}$
I_0 := Closure([S' → .S, $])
repeat until C can't be completed:
For each set of edges $I_i \in C$ and each symbol $X \in V \cup T$
complete I_{i+1} by GOTO-EDGES(I_i, X).

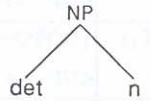
Calculation of the cell entries:
a) ACTION-cells
For I := 0 to n:
If $[A \rightarrow \alpha.a\beta, b] \in I_i$ and GOTO-EDGES(I_i, a) = I_j
Then ACTION(i, a) = shift j
If $[A \rightarrow \alpha., a] \in I_i$
Then ACTION(i, a) = reduce p and p is index of rule $A \rightarrow \alpha$.
If $[S' \rightarrow S., \$] \in I_i$
Then ACTION(i, $) = accept

b) GOTO-cells
For all I_i and $X \in V \cup T$:
If GOTO-EDGES(I_i, X) = I_j
Then GOTO(i, X) = j
Set all empty cells to ERROR

3 Die Verwendung komplexer Symbole

Jeder Parser, der die Wohlgeformtheit von Sätzen zu prüfen hat, muß in irgendeiner Weise die Kontrollmechanismen der Grammatik reflektieren, die diese Wohlgeformtheit garantieren, d.h. wenn die Grammatik komplexe Symbole verwendet, so muß dem Parser ein analoger (oder einfacher: der gleiche) Formalismus zu Verfügung stehen, um die gleichen oder ähnliche Strukturen erkennen und erzeugen zu können. Die Unifikation in der LFG (lexical functional grammar) ist ein solcher Kontrollmechanismus: Wenn zwei Merkmalmengen *A* und *B* unifiziert ($A \sqcup B$) werden sollen, ist dies nur möglich, wenn die Werte für alle Merkmale $f \in A$ und $f \in B$ gleich sind. Verschiedene Merkmalnamen hingegen werden einfach vereint.
Ein Beispiel macht dies deutlich:

det	= {<plu, +>, <def, +>}
n	= {<plu, +>, <hum, +>}
det ⊔ *n*	= {<plu, +>, <hum, +>, <det, +>}



Für den Teilbaum heißt dies, daß bei der Unifikation NP = det ⊔ n, *det* und *n* im Plural stehen müssen. Ist dies nicht möglich, wird der komplette Satz zurückgewiesen. Erinnern

wir uns, wie der Tomita-Parser seine Stacks reduziert: Befinden sich die zwei Symbole *A* und *B* an oberster Stelle des Stacks und es existiert eine Regel $C \rightarrow A\ B$, dann entferne *A* und *B* und push' *C*. An dieser Stelle müssen wir komplexe Symbole berücksichtigen. Wenn *A*, *B* etc. komplexe Symbole darstellen, dann dürfen sie nur dann vom Stack entfernt werden, wenn sie zu *C* unifiziert werden können. Das heißt, daß der Stack oder zugehörige Datenstrukturen in der Lage sein müssen, die Merkmale ihrer Elemente aufzunehmen.

Somit ist am Tomita-Algorithmus selbst keine Änderung notwendig, um komplexe Symbole zu verkraften, lediglich die Verwaltung des Stacks und der Strukturen bedarf einer Anpassung an die jetzt neu hinzugekommenen Merkmalbündel.

4 Beispiel

Betrachten wir einmal die folgenden (sehr einfachen) ID-relations zur Generierung von (sehr einfachen) deutschen Sätzen:

G: (1) S → NP, VP
(2) NP → pro
(3) VP → v, NP

Diese Grammatik generiert *Jeder kennt ihn. Ihn kennt jeder. Kennt ihn jeder?* aber ebenso die falschen Sätze **Kennt jeder ihn?* und **Ihn jeder kennt* wenngleich dies nur erlaubt sein darf, wenn die Baumstruktur crossing-over gestattet. Einige dieser Übergenerierungen werden wir später eliminieren. Darüberhinaus wird **Jeder ihn kennt* ebenfalls generiert, doch als Tribut an die Einfachheit der Beispielgrammatik sind wir bereit, es zu ignorieren.

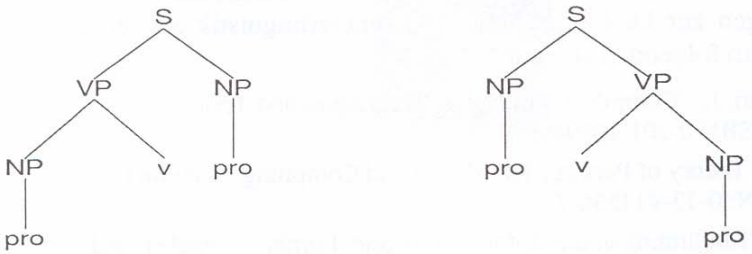
Der modifizierte Tabellengenerator erzeugt aus obiger Grammatik zunächst folgende Kanten:

I_0	<S' → .S, $>, <S → .NP VP, $> <S → .VP NP, $> <NP → .pro, v> <NP → .pro, pro> <VP → .v NP, pro> <VP → .NP v, pro>	I_1	[GOTO-EDGES(I_0, S)] <S' → S., $>	I_2	[GOTO-EDGES(I_0, NP)] <S → NP.VP, $> <VP → NP.v, pro> <VP → .v NP, $> <VP → .NP v, $> <NP → .pro, v>
I_3	[GOTO-EDGES(I_0, VP)] <S → VP.NP, $> <NP → .pro, $>	I_4	[GOTO-EDGES(I_0, pro)] <NP → pro., v> <NP → pro., pro>	I_5	[GOTO-EDGES(I_0, v)] <VP → v.NP, pro> <NP → .pro, pro>
I_6	[GOTO-EDGES(I_2, VP)] <S → NP VP., $>	I_7	[GOTO-EDGES(I_2, v)] <VP → NP v., pro> <VP → v.NP, $> <NP → .pro, $>	I_8	[GOTO-EDGES(I_2, NP)] <VP → NP.v, $>
I_9	[GOTO-EDGES(I_2, pro)] <NP → pro., v>	I_{10}	[GOTO-EDGES(I_3, NP)] <S → VP NP., $>	I_{11}	[GOTO-EDGES(I_3, pro)] <NP → pro., $>
I_{12}	[GOTO-EDGES(I_5, NP)] <VP → v NP., pro>	I_{13}	[GOTO-EDGES(I_5, pro)] <NP → pro., pro>	I_{14}	[GOTO-EDGES(I_7, NP)] <VP → v NP., $>
I_{15}	[GOTO-EDGES(I_7, pro)] <NP → pro., $>	I_{16}	[GOTO-EDGES(I_8, v)] <VP → NP v., $>		

Das führt zu folgender Parsing-Tabelle:

	S	NP	VP	pro	v	$		S	NP	VP	pro	v
0	sh1	sh2	sh3	sh4	sh5			1	2	3	4	5
1						acc						
2		sh8	sh6	sh9	sh7				8	6	9	7
3		sh10		sh11					10		11	
4				re2	re2							
5		sh12		sh13					12		13	
6						re1						
7		sh14		sh15 re3					14		15	
8					sh16							16
9					re2							
10						re1						
11						re2						
12				re3								
13				re2								
14						re3						
15						re2						
16						re3						

Unter Annahme eines geeigneten Lexikons (ohne Merkmale!), erhalten wir durch den Parsing-Prozeß folgende Strukturen für den Satz *Jeder kennt ihn*:



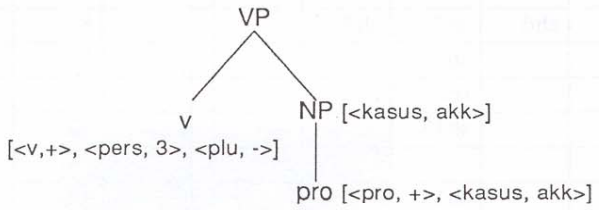
Während nur der rechte Ableitungsbaum für den gegebenen Satz korrekt ist, wäre der linke Baum eine korrekte Ableitung für *Ihn kennt jeder*. Dies ist das Resultat einer unerwünschten (und falschen) Reduktion des Satz-Subjektes mit dem Verb der VP. Im umgekehrten Fall würden für den Satz *Ihn kennt jeder* die gleichen Strukturen erzeugt werden und der rechte Baum wäre falsch - aus den gleichen Gründen. Um dies zu verhindern, werden wir von Merkmalen wie im untenstehenden Lexikon Gebrauch machen:

L:	kennt:	{<v, +>, <pers, 3>, <plu, ->}
	ihn:	{<pro, +>, <case, akk>}
	jeder:	{<pro, +>, <case, nom>}

Wir modifizieren die obige Grammatik wie folgt:

G:	(1) S → NP[<case, nom>], VP
	(2) NP → pro
	(3) VP → v, NP[<case, akk>]

Tatsächlich ist es irrelevant, ob die Merkmale im Lexikon oder in der Grammatik instanziiert werden. Das beeinflußt die Generierung der Parsing-Tabelle in keiner Weise. Wir sollten darauf hinweisen, daß es elegantere Wege gibt, für eine VP ein Objekt zu erzwingen, doch im Moment reicht es völlig, wenn der Parser die Merkmale in einer geeigneten Datenstruktur festhält. Im Beispiel unten ist die Wahl von *ihn* statt *jeder* notwendig, soll die Unifikation erfolgreich sein.



Nun produziert der Parser für *Jeder kennt ihn* und *Ihn kennt jeder* die korrekten Strukturen und weist den Satz **Ihn jeder kennt* als ungrammatisch zurück.

Bisher sprachen wir nicht über notwendige Mechanismen des feature parsings. Tatsächlich sind die Mechanismen zur Behandlung von Merkmalen in zum Beispiel LFG und GPSG so verschieden, daß keine Chance besteht, diese Grammatiken in gleicher Weise zu behandeln. Hingegen ist im obigen Beispiel, wie schon angedeutet, die Merkmalmenge der Elternknoten die Unifikation der Merkmalmengen seiner Kinder. Das ist nicht immer möglich. So werden NP und VP im obigen Beispiel aufgrund ihrer verschiedenen Kasus niemals zu S unifiziert werden können. Ein Vorschlag: Wir können die Zuweisung der Funktionsnamen in LFG als eine Zuweisung von Namen von Merkmalsmengen auffassen (und müssen hierfür die reduce-Operation des Tomita-Algorithmus angemessen modifizieren), so daß die verschiedenen Merkmalswerte von NP und VP deshalb unifizieren, weil sie unter verschiedenen Namen verwaltet werden, z.B. *subject* und *object*. Aber es sind noch viele andere Lösungen denkbar.

Die Grundidee der in diesem Artikel beschriebenen Algorithmenmodifikationen entstand im Rahmen meiner Vorbereitungen zur Diplomprüfung in Computerlinguistik an der Universität Koblenz. Zur Verwendung kam folgende Literatur:

[1] Aho, A. / Sethi, R. / Ullman, J.: "Compilers. Principles, Techniques and Tools" Addison Wesley 1986, ISBN 0-201-10194-7

[2] Aho, A. / Ullman, J.: "The Theory of Parsing, Translating and Compiling. Volume I" Prentice-Hall 1972, ISBN: 0-13-914556-7

[3] Hopcroft, J. / Ullman, J.: "Einführung in die Automatentheorie, Formale Sprachen und Komplexitätstheorie"; Addison Wesley 1990, ISBN: 3-89319-181

[4] Naumann, S. / Langer, H.: "Parsing" ; Teubner 1994, ISBN: 3-519-02139-0

[5] Tomita, M.: "Efficient Parsing for Natural Language"; Kluwer Academic Publishers 1986, ISBN: 0-89838-202-5

5 Glossar

5.1 ID/LP Grammatik:

Einfache Phrasenstrukturgrammatiken (PSG) verknüpfen zwei syntaktische Relationen meist so, daß ihre explizite Generalisierung unmöglich wird. Darum werden Relationen in ID/LP-Grammatiken getrennt formuliert: Die immediate-dominance-Regeln (ID) beschreiben die Dominanzbeziehungen zwischen linker und rechter Regelseite (z.B. *S* wird ersetzt durch *NP* und *VP*). Die linear-precedence-Regeln (LP) legen dabei fest, in welcher Reihenfolge. Liegt keine explizite Regel vor, so ist die Reihenfolge beliebig. Die **LP-Expansion** einer ID/LP-Regel ist dann die Menge aller aus ihr konstruierbaren PSG-Regeln. Aus der ID-Regel $C \rightarrow A, B$ (man beachte das Komma) ohne weitere LP-Regeln läßt sich so $C \rightarrow A\ B$ und $C \rightarrow B\ A$ konstruieren.

5.2 Lexical Functional Grammar (LFG):
Sie gehört zur Klasse der unifikationsbasierten Grammatiken, d.h. die Symbole der Phrasen sind nicht einfach Bezeichner einer substanzlosen Struktur (z.B. S für eine Satzstruktur), sondern Bezeichner einer Attributmenge und mithin komplexe Symbole. Jedoch sind diese Attribute nicht beliebig, sondern dienen der funktionalen Beschreibung der Satzkonstituenten. Das Subjekt des Satzes *Peter ißt eine Tomate* wird dann wie folgt notiert:

Subjekt	Präd: Peter Genus: MSK Num: SG Kasus: NOM Hum: +

5.3 Kontextfreie Grammatik (efg):

Diese Klasse stellt all' jene Grammatiken dar, die nur solche Regeln enthalten, für deren Anwendung keinerlei Kontextbedingungen gelten, d.h. auf deren linke Regelseiten keine Terminalsymbole (Wörter) vorkommen. Die PSG Z.B. ist somit kontextfrei.

5.4 LR(k)-Parser:

Diese Parserklasse arbeitet Eingaben (Sätze) von links nach rechts ab und generiert dabei Rechtsreduktionen. Das *k* gibt die Anzahl der noch nicht bearbeiteten Symbole der Eingabe an, die betrachtet werden müssen, um Entscheidungen für den nächsten Schritt treffen zu können. Diesen Entscheidungen liegt eine Parsing-Tabelle zugrunde, in der zu jedem Zustand und gegebenem nächsten Symbol steht, welche Aktion und welcher Folgezustand zu wählen ist. Sie wird durch einen Tabellengenerator für jede Grammatik einmal erzeugt. Ein Parsingprozeß sieht prinzipiell so aus: Der Parser prüft, ob das oberste Symbol seines Stacks (der zu Beginn bis auf ein Startsymbol leer ist) und das aktuelle Symbol der Eingabe der rechten Seite einer Regel entsprechen. Ist dies der Fall, so werden sie entfernt (reduce) und durch die linke Regelseite ersetzt (z.B. ersetze *V* und *NP* bei vorliegender Regel *VP* ---) *V NP* durch *VP).* Ist dies nicht möglich, so legt der Parser das aktuelle Symbol auf den Stack (shift), wählt das nächst rechte Symbol als das aktuelle aus und fahrt wie oben fort. Ist kein shift oder reduce möglich, ist entweder der Stack leer und die Eingabe erfolgreich abgearbeitet, oder es befinden sich noch Symbole im Stack und der Parser bricht die Bearbeitung mit einem Fehler ab.

MASCHINEN LERNEN SPRACHE - DIE COMPUTERLINGUISTIK

Programm Bayern 2 Dienstag 6. Juni 95, 19:30 Uhr

Susanne Tölke, bearbeitet von Prof. Wolfgang Hoeppner

Zuspielung

Guten Tag, hier ist Lisa, das Sprachdialogsystem der Lufthansa. Wenn Sie eine grundlegende Auskunft wünschen, sagen Sie nach dem Pfeifton "null", wenn Sie Lisa nur gelegentlich benützen, sagen Sie "eins", wenn Sie Profi sind, sagen Sie "zwei"…

Sprecherin

Die automatische Flugauskunft der Lufthansa. Wenn alles gutgeht, merkt man gar nicht, daß man mit einem Computer gesprochen hat.

Sprecher

Die natürliche Sprache wird noch nicht lange als Interaktionsform zwischen Mensch und Computer eingesetzt. Wer mit dem Computer arbeiten wollte, mußte zunächst die Sprache der Maschine lernen. Eine Aufgabenstellung mußte in dieser Sprache formuliert werden, um vom Computer erkannt zu werden. Für die verschiedenen Arbeitsbereiche, in denen Software produziert wird, entwickelte man jeweils eigene Sprachen, ja sogar für einzelne Bereiche wie Datenbanksysteme oder Betriebssysteme existieren zahlreiche Systemtypen mit jeweils speziellen Anfragesprachen. Die Menschen dagegen, die doch auch in sehr unterschiedlichen Arbeitsabläufen und Problembereichen zusammenarbeiten, haben nur eine einzige Sprache. Diese Sprache, bei uns das Deutsche, wird von den Linguisten als "natürliche Sprache" bezeichnet, im Gegensatz zu den "formalen Sprachen", die für die Software-Anwendungen geschaffen werden.

Sprecherin

Die natürlichen Sprachen haben den Vorteil, daß sie universell einsetzbar sind und von allen Menschen - aufgrund ihrer Sozialisation - beherrscht werden. Warum also nicht versuchen, natürliche Sprache als Interaktionsform zwischen Mensch und Computer einzuführen? Die Computerlinguistik, auch linguistische Informatik genannt, befaßt sich mit der Frage: Lassen sich Computer so programmieren, daß sie natürlichsprachliche Aufgabenstellungen verstehen?

Sprecher

Dahinter steht der Wunsch, die Ausgaben der Software-Systeme für den Benutzer leichter verständlich zu machen. Dazu Reinhard Knerser, der die automatische Zugauskunft betreut:

Zuspielung (Knerser)

Wir haben uns bemüht, daß nicht der Benutzer sich an den Computer anpassen muß, sondern daß sich der Computer an den Benutzer anpaßt. Das heißt, der Benutzer sollte ganz normal mit dem Computer kommunizieren können, wie er es auch mit einem Bahnbeamten machen würde. Es ist wichtig, Korrekturen anzubringen, wie man das auch in einer gestörten Unterhaltung machen würde. Der Computer versteht das, wenn man sagt: Nein, ich wollte nicht morgen fahren, ich wollte am Sonntag fahren. Wir haben beobachtet, daß 80% der Anrufer erfolgreich eine Auskunft kriegen.

Sprecherin

Die fehlenden 20% gehen zum kleineren Teil darauf zurück, daß nach Bahnhöfen gefragt wurde, die noch nicht im Programm sind, zum größeren Teil jedoch auf Hörfehler des Computers.

Sprecherin

Wer "Regensburg" mit zu geschlossenem "e" spricht, wird vielleicht vom Computer die Anschlüsse für den Bahnhof "Dresden" bekommen, wer bei "München" zu sehr das "ch" betont, wird unter Umständen auf die Stadt "Aachen" verwiesen. Der Computer kann

eben nur auf Schallsignale reagieren, und die klingen sehr unterschiedlich, je nachdem ob der Benutzer deutlich spricht oder nuschelt, eine hohe oder tiefe Stimme hat und aus Ostfriesland kommt oder aus Bayern.

Sprecher
"Mangelnde Spracherkennung" nennen das die Sprachwissenschaftler, die dem Computer hören und sprechen beibringen wollen. Dabei ist die Variabilität der Aussprache noch das kleinere Problem. Das größere ist die fließende Rede. Dazu Wolfgang Hoeppner, Professor für Computerlinguistik an der Universität Duisburg:

Zuspielung (Hoeppner)
"Das ist gar nicht so einfach und liegt daran, daß in gesprochener Sprache zwischen den einzelnen Wörtern keine Pausen sind. Man denkt immer, da sind Pausen drin, es sind manchmal Pausen drin, gerade habe ich eine gemacht, aber die sind dann nicht bedeutungstragend. Und es ist sehr schwierig, aus diesen Schallsignalen die bedeutungsvollen Einheiten 'rauszukriegen. Bei geschriebener Sprache hat man die Leerzeichen und das trennt die Wörter schon mal."

Sprecher
Die Spracherkennung ist für den Computer deshalb problematisch, weil er nicht über das enorme Wissen verfügt, das dem menschlichen Hörer ermöglicht, auch unverständliche Teile in einen sinngemäßen Gesamtzusammenhang einzubetten. Die Sinnzuordnung fehlt der Maschine. Das gilt für die undeutlichen Wörter ebenso wie für die mehrdeutigen. Solche Probleme haben vor allem Diktiercomputer. Sie müssen bei mehrdeutigen Aussagen entscheiden, was gemeint ist, zum Beispiel bei dieser:

Sprecherin
Der Gefangene floh

Sprecher
Das kann im Zusammenhang heißen:

Sprecherin
Das Tor stand offen und der Gefangene floh.

Sprecher
Es kann aber auch heißen:

Sprecherin
Schau mal, in der Streichholzschachtel sitzt der gefangene Floh!

Sprecher
Der Mensch versteht den Sinnzusammenhang des Satzes und weiß Bescheid. Der Computer muß auf statistischer Basis entscheiden: Welche Lösung ist die wahrscheinlichere?

Sprecherin
Immerhin gibt es bereits Systeme, die eine Folge von drei Wörtern überblicken und damit viele Mehrdeutigkeiten unterscheiden können. Ludwig Hitzenberger, Dozent für Computerlinguistik an der Universität Regensburg:

Zuspielung (Hitzenberger)
"Die Mehrdeutigkeiten werden durch ein Sprachmodell aufgelöst. Das sind statistische Modelle, die über sehr großen Wortschätzen trainiert werden. Die Auflösung von Homophonen und Homonymen ist kein Problem. Wenn ich sage: 'Das ist der springende Punkt Punkr, kann das System ohne weiteres erkennen, daß das eine das Nomen Punkt ist und das andere das Satzzeichen Punkt."

Sprecherin
An der Verbesserung der Spracherkennung wird laufend gearbeitet, denn der Bedarf an solchen Computern ist groß. Wolfgang Hoeppner:

Zuspielung (Hoeppner)
"Das versucht man heute für Arbeitsplätze, an denen man Hände nicht einsetzen kann oder in denen es dunkel ist. Röntgenärzte sind ein beliebtes Problemfeld, wo man sagt, die sind in einem dunklen Raum, die sollen auf die Röntgenbilder gucken und ihre Diagnose hineinsprechen, und das soll als Arztbericht weitergereicht werden an den nachbehandelnden Arzt. Das wäre so eine typische Anwendung."

Sprecher
Eine weitere Anwendung ist der Computer als Hilfsgerät für Behinderte. Es gibt Sprachcomputer, die nur auf eine Stimme reagieren, und gerade diese Systeme sind weniger störanfällig - eine ideale Unterstützung für behinderte Benutzer.

Sprecherin
Man sollte sich also vor dem Erwerb eines Spracherkenners genau überlegen, auf welche Funktionen man besonderen Wert legt. Ludwig Hitzenberger:

Zuspielung (Hitzenberger)
"Wir sind im Augenblick in der Lage, daß wir kommerziell verfügbare Spracherkenner ha

ben, die nicht mal so sehr teuer sind: Größenordnung 2500 Mark, und diese Spracherkenner verstehen einen Sprachschatz von 25-30.000 Wortformen, und damit ist es möglich, enge Anwendungen, bei denen der Wortschatz nicht allzu üppig ist, zum Beispiel Radiologiebericht, mit einer sehr geringen Fehlerquote zu diktieren. Es gibt verschiedene Formen der Spracherkenner. Die kurze Pause nach Wörtern, das isolierte Sprechen, erleichtert die Spracherkennung ungemein und erlaubt dadurch größere Wortschätze. Bei kontinuierlicher Sprache ist der Wortschatz reduzierter, aber eben kontinuierliche Sprache möglich. Ein dritter Parameter ist die Sprecherabhängigkeit. Systeme, die sprecherunabhängig arbeiten, haben wesentlich größere Probleme, etwas zu erkennen, als wenn das System auf einen bestimmten Sprecher normieren kann. Jeder Mensch spricht unterschiedlich. Undeutliches Sprechen ist kein großes Problem, solange man immer undeutlich spricht. Wenn das System auf einen Sprecher normiert ist und der eine schlampige Aussprache hat, solange der das immer gleich macht, ist es erkennbar. Es ist gerade für den Behindertenbereich sehr interessant."

Sprecherin

Der dritte zukunftsträchtige Bereich - neben den Auskunfts- und den Diktiersystemen - ist das Dolmetschen. Seit Mitte der achtziger Jahre sind elektronische Wörterbücher auf dem Markt, und es wird nicht mehr lange dauern, bis sie auch mit einem Spracherkenner ausgerüstet sind und auf Ansprache reagieren. Willy Martin, Professor für Lexikographie an der Universität Amsterdam, arbeitet an solchen elektronischen Wörterbüchern:

Zuspielung (Martin)

"Das ist für die Europäische Union sehr wichtig, daß die Infrastrukturen gut sind. Infrastrukturen sind nicht nur Wege, Straßen, sondern auch Sprachen. Alle Sprachen, nicht nur die großen Sprachen. Einheit bedeutet auch Vielheit. "

Sprecher

Die automatische Sprachverarbeitung gilt als vielversprechende Technologie der Zukunft. Weltweit wird auf diesem Gebiet gearbeitet, vor allem in Japan, Amerika und Europa, doch bisher hat sich noch keine Vormachtstellung gebildet. Deutschland hat bei der Entwicklung dieser Spitzentechnologie noch alle Chancen.

Sprecherin

Das Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie, im Volksmund "Zukunftsministerium" genannt, hat für den Zeitraum von 1993 *bis* 1996 66,4 *Millionen Mark für die Entwicklung des Dolmetsch-Computers "Verbmobil" zur Verfügung gestellt. Rund* 34 *Millionen Mark zahlen die Industrieunternehmen, die mit ihren Forschungsabteilungen auch an diesem Projekt beteiligt sind. Die wissenschaftliche Leitung obliegt dem Deutschen Forschungszentrum für Künstliche Intelligenz in Kaiserslautern, weitere* 19 *Hochschulen sind angegliedert. Was soll die Maschine leisten können?*

Sprecher

Das tragbare Gerät "Verbmobi1" soll bei Gesprächen zwischen fremdsprachlichen Partnern eingesetzt werden. Beide Partner müssen Englisch beherrschen, können aber bei ausgefallenen Redensarten oder komplexen Satzkonstruktionen ihre Muttersprache benützen - das Verbmobil übersetzt dann ins Englische. Als Eingabesprachen für den Forschungsprototyp sind Deutsch und Japanisch ausgewählt worden, der modulare Ansatz ist jedoch auf andere Eingabesprachen übertragbar.

Sprecherin

1m Februar wurde das erste Zwischen ziel des Projekts erreicht, der sogenannte "Demonstrator". Er kann einfache, grammatikalisch korrekte Sätze vom Japanischen ins Englische und vom Deutschen ins Englische übersetzen. Ende 1996 *soll Verbmobil auch mit einigen Elementen unvollständiger Sprache umgehen können. Die letzte Phase, in etwa acht Jahren, soll ein Computer sein, der flüssig gesprochene, grammatikalisch korrekte Sprache übersetzen kann und einen Wortschatz von* 3000 *Wörtern beherrscht. Mit diesem Umfang kann man, so die Computerlinguisten, schon* 90 % *des Alltags bestreiten.*

Sprecher

Damit hätte Verbmobil zwei oder drei Schwierigkeitsgrade der Spracherkennung bewältigt: Am einfachsten ist die sogenannte Einzelworterkennung, also das Sprechen mit künstlicher Pause nach jedem Wort. Die zweite Stufe ist die kontinuierliche, fließende

Sprache, die phonetisch und grammatikalisch korrekt ist. Am schwierigsten ist die Erkennung der Spontansprache mit all ihren abgebrochenen Sätzen, Ausrufen und Einschüben, so wie sie im freien Dialog vorkommt.

Sprecherin
Es tun sich aber auch jene Übersetzungscomputer schwer, die ohne Spracherkenner arbeiten, die geschriebene Texte verarbeiten. Die natürliche Sprache hat auch hier ihre Tücken. Sie hat sich bisher der vollständigen formalen Analyse entzogen. Und ob es jemals ein System geben wird, das es an Flexibilität mit der natürlichen Sprache aufnehmen kann, wird von einigen Wissenschaftlern bezweifelt. Einige ziehen es auch vor, sich lieber auf den engen, anwendungsorientierten Bereich zu konzentrieren, weil hier schnelle, befriedigende Erfolge winken. Sie gestalten lieber Software, die auf einen bestimmten Bedarf zugeschnitten ist. Damit sind sie eigentlich mehr Informatiker als Sprachwissenschaftler. Einer von ihnen ist Jürgen Krause, bislang Professor für linguistische Informationswissenschaften in Regensburg, seit 1. *Mai Professor für Informatik an der Universität Koblenz.*

Zuspielung (Krause)
"Vor 20 Jahren war die natürliche Sprache die einzige Möglichkeit, Computer benutzerfreundlich zu gestalten, und ab den 80er Jahren hat man dann gesehen, daß es da auch andere Möglichkeiten gibt, mit der Maus und dem Bildchen, und insofern ist das von der Sache her eine natürliche Entwicklung. Mir ist es in der Forschung immer darum gegangen, benutzerfreundliche Software zu gestalten, und da hat sich das Paradigma geändert. Die natürliche Sprache ist nicht mehr die vorherrschende Grundidee und die graphischen Oberflächen haben einen Vorsprung. Die Natürlichsprachlichkeit ist demgegenüber stark zurückgesetzt. Ich glaube, wenn die Computerlinguistik rein in der Sprachwissenschaft bleiben würde, sich verstehen würde als formale Linguistik, daß es dann eigentlich ein absterbender Ast ist. Ich glaube wirklich, die Chance der Computerlinguistik liegt in der Verbindung zur Informatik und zur Software. Ich seh das auch ganz deutlich an den Studienabgängern. Wir haben in den fünfzehn Jahren, in denen ich das Fach vertrete, keinen einzigen Arbeitslosen gehabt. Wir haben immer mehr offene Stellen gemeldet, als Kandidaten da waren, und das zeigt nach meiner Ansicht: Man sollte die Computerlinguistik in enger Verbindung zur Informatik sehen und auch studieren."

Sprecherin
Wer als Informatiker im Bereich "Sprache und Computer" arbeitet, wird den Zugang über die formale Sprache des Computers suchen. Wer als Linguist in diesem Bereich arbeitet, wird den Zugang über die natürliche, das heißt die historisch gewachsene und sich ständig wandelnde menschliche Sprache suchen. Jürgen Krause ist eher skeptisch, was den natürlichsprachlichen Zugang betrifft.

Zuspielung (Krause)
"Ich glaube, daß für viele Datenbankbereiche natürlichsprachliche Zugangswege möglich sind, daß es aber bessere gibt. Der Benutzer sieht, daß er mit den graphischen Möglichkeiten besser zurechtkommt. Es gibt Bereiche, wo die Natürlichsprachlichkeit eingesetzt werden kann, aber das sind Nischen. Die vorherrschende Lehre sind die graphischen Oberflächen."

Sprecherin
Den Siegeszug der graphischen Oberfläche führt Jürgen Krause auf zwei Gründe zurück: Zum einen glaubt er nicht an den interdisziplinären Ansatz, der für die Erforschung des Mensch-Maschine Dialogs nötig ist. Die Versuche in dieser Richtung hätten doch bisher recht wenig gebracht:

Zuspielung (Krause)
"Man nimmt einen Informatiker, einen Kognitionswissenschaftler, einen Psychologen und einen Linguisten und die sperrt man in ein Zimmer und die zwingt man, miteinander zu reden und die sollen dann ein Frage-Antwort-System machen. Und wenn sie das tun, ich hab auch schon in solchen Gruppen gearbeitet, dann werden sie schnell feststellen, daß das nicht klappt, weil die Leute nicht kommunizieren können. Es gibt Professoren, die sind Linguisten, und es gibt Professoren, die sind Informatiker, und vom armen Studenten verlangt man jetzt, er macht ein Zweifachstudium - Informatiker und Linguist. Das sind diese interdisziplinären Studiengänge, ich glaube nicht, daß das der richtige Weg ist."

Sprecherin
Das zweite Hemmnis für die Entwicklung des Mensch-Maschine-Dialogs sieht Krause in dem Umstand, daß Menschen sich anders

verhalten, wenn sie mit einem Computer reden.

Zuspielung (Krause)

"Uns fiel sehr schnell auf, daß die Leute, wenn sie mit dem Computer sprechen, anders sprechen, als wenn sie mit dem Menschen sprechen. Das ist die sogenannte Computer-Talk-Hypothese, und das war für uns eine der Erklärungen dafür, daß solche Systeme nicht laufen, wenn Linguisten die entwickeln."

Sprecherin

Um die Computer-Talk-Hypothese zu beweisen, setzten die Wissenschaftler ihre Probanden vor ein Mikrophon und ließen sie verschiedene Auskunftssysteme anrufen.

Zuspielung (Krause)

"In dem einen Fall haben wir gesagt, du redest jetzt mit einem Bahnbeamten, und in dem andern Fall haben wir gesagt, du redest mit einem hochentwickelten Computersystem. Und wenn sie diese zwei Dinge vergleichen, dann stellen sie fest, daß die Leute ganz andere Ausdrücke benützen, andere Syntaxstrukturen, also ganz anders sprechen mit dem System, wo sie glauben, da ist eine Maschine dahinter und mit dem System, wo sie glauben, da ist ein Bahnbeamter dahinter."

Sprecherin

Auch eine Auskunft für Eltern schulpflichtiger Kinder wurde installiert, einmal - angeblich - mit einem Beratungslehrer, einmal mit einem Computer. Auch hier unterschieden sich die Fragen.

Zuspielung (Krause)

"Es wurde zum Beispiel der Satz gebildet: "Welche Deutschnote in Quarta hat wieviele Schüler?" Das ist ein Satz, der unakzeptabel ist, im Deutschen, den gibt's gar nicht. Und was eigentlich gewollt war: "Wieviele Schüler haben welche Deutschnote in Quarta?" Da wurden die beiden Substantive umgedreht. Und wenn man sich anschaut, was der Grund ist, dann stellt man fest, die wurden umgedreht, weil das Ergebnis, das sie durch den Computer bekamen, das hat die Deutschnote in der Tabelle in der Spalte 1 gehabt. Das haben die Leute gesehen und da wollten sie's dem Computer leichter machen, und darum haben sie unakzeptable Sätze gebildet. Das sind die sogenannten Computer- Talk-Eigenschaften. Die wird kein Linguist untersuchen, weil der an der natürlichen Sprache interessiert ist."

Sprecherin

Die Computer-Talk-Eigenschaften sind also so sieht es Jürgen Krause - ein zusätzliches Hemmnis bei Entwicklung eines funktionierenden Mensch-Maschine-Dialogs. Er sieht hier keine große Zukunft. Allerdings gesteht er der Grundlagenforschung im Bereich der natürlichen Sprache eine grundsätzliche Bedeutung zu:

Zuspielung (Krause)

"Wir brauchen beides: Sowohl die Grundlagenforschung von der theoretischen Seite, von der Linguistik her, mit dem Ziel, einen Computer zu konstruieren, der die natürliche Sprache genauso wie der Mensch beherrscht, aber wir müssen auf der anderen Seite sehen, daß die Unmöglichkeit, dieses Ziel zu erreichen, kein Hinderungsgrund ist, gute Software zu gestalten. Es gibt sehr viele Bereiche, wo man mit einem eingeschränkten Wissen, eingeschränkten Wortschatz, eingeschränkten syntaktischen Strukturen durchaus Erfolg haben kann, und das ist ja das Entscheidende. Wenn ich einen Bereich hab wie bestimmte Datenbanken über das Personal einer Firma oder Stücklisten usw und ich kann alle Fragen, die an diese Datenbank gestellt werden, linguistisch beherrschen, dann genügt das durchaus, um Software zu verkaufen. Anwendungsorientierte Computerlinguistik mißt sich daran, ob ich Software konstruieren kann, die funktioniert und deshalb auch gekauft wird. Und das ist für die Studenten der Markt. In diesen Bereichen werden die ihre Stellen bekommen. Auf der anderen Seite ist völlig klar, daß auf den Universitäten der theoretische Bereich auch da sein muß. Man muß sehen, daß das ein sehr kleiner Bereich ist. Das werden immer sehr wenig Leute sein."

Sprecherin

Einer dieser wenigen ist Roland Hausser, Professor für linguistische Informatik an der Universität Erlangen. Er verficht hartnäckig die Idee, daß sich die Vielfalt der natürlichen Sprache eines Tages durch ein formales Modell darstellen lassen wird:

Zuspielung (Hausser)

"In der Informatik unterscheidet man zwischen smart solutions, schlauen Lösungen. und solid solutions, soliden Lösungen. Solid solutions findet man im Bereich der Grundlagenforschung, dort versucht man ein Problem in seiner Gänze zu verstehen und zu lösen. Smart solutions findet man in den Anwendungen, wo der Benutzer mit einer 80%igen Lösung zufrieden ist. Verständlicherweise hat sich eine Art Richtungskampf zwischen den Vertretern der smart solution und der solid solution eingestellt. Die Vertreter der smart solution verweisen auf die schnelle kommerzielle Einsetzbarkeit ihrer

Produkte, wobei sie manchmal nicht erwähnen, daß eine 80%ige smart solution aus einer Reihe von Gründen grundsätzlich nicht auf eine 90%ige oder gar 95%ige Lösung erweitert werden kann."

Sprecher

Für die anwendungsorientierte smart solution spricht natürlich, daß sie umgehend kommerziell verwertbar ist, die grundlagenorientierte solid solution kostet zunächst einmal nur Geld. Lohnt sich das?

Zuspielung (Hausser)

"Die solid solution im Bereich der Grundlagenforschung ist natürlich sehr aufwendig. Wenn eine solche solid solution, also im Bereich der Sprachwissenschaft zum Beispiel ein vollständiges Lexikon des Deutschen mit einer zugehörigen Morphologie aber erst einmal fertiggestellt ist, dann kann es in kommerziellen Anwendungen als sogenanntes "off-the-shelfproduct", also ein Produkt, das man vom Regal nimmt, eingesetzt werden und bietet dann eine sehr gute Grundlage für Lösungen, die in den Bereich von 95-99% hineingehen."

Sprecherin

Sogar für die Grammatik, den komplexesten und schwierigsten Bereich der Linguistik, sieht Roland Hausser einen Durchbruch voraus:

Zuspielung (Hausser)

"Nach meiner Überzeugung sind die Grammatiken nur deshalb so kompliziert, weil sie von Voraussetzungen ausgehen, die die empirischen Gegebenheiten noch nicht so richtig beschreiben. So wie im Mittelalter die Astronomen sehr komplizierte Systeme hatten, um die Planetenbahnen zu beschreiben, bis Kepler die Ellipse als grundlegendes Prinzip fand, so haben wir heute sehr komplizierte Grammatiken, aber ich bin der Überzeugung, daß sich in einigen Jahren die Grammatiken sehr vereinfacht haben werden."

Sprecher

So wie es heute schon Rechtschreib-Checker gibt, würde es dann Grammatik-Checker geben - ein Fortschritt, den so mancher Computerbenutzer begrüßen würde. Noch interessanter wäre es freilich auf der nächsten Ebene, der semantischen. Der Computer könnte die syntaktische Analyse auch inhaltlich interpretieren. ER könnte für uns Zeitung lesen und das Wichtige heraussortieren.

Zuspielung (Hausser)

"Da sehe ich die wirklichen Anwendungsgebiete der Computerlinguistik. Es heißt, daß sich die gedruckte Information auf der Erde alle zehn Jahre verdoppelt, so daß es für den einzelnen Menschen gar nicht möglich ist, das alles zu lesen, und hier kann mit einer automatischen Inhaltsanalyse Abhilfe geschaffen werden. Das ist noch ein Zukunftstraum, aber ich denke, in 20 Jahren wird man auf diesem Gebiet wesentlich weiter sein und es wird ein Wandel stattfinden wie der vom Pferdekarren zum Auto. In diesem Zusammenhang wird meist eingewendet, daß das zu einem Verlust führen würde, aber genauso wenig wie das Auto uns den Verlust unsrer Gehfähigkeit gebracht hat, genausowenig werden diese automatischen Systeme zur Inhaltsanalyse unsere intellektuellen Fähigkeiten auf dem Gebiet des Lesens und Zuhörens mindern."

Sprecherin

Ein Computer, der für uns das Lesen und Exerpieren übernimmt - ein ehrgeiziges Projekt. Es ist ja immer noch ungeklärt, wie wir mit unseren sprachlichen Fähigkeiten umgehen.

Sprecher

Die einen versuchen, durch möglichst umfassende empirische Datensammlungen an das Geheimnis heranzukommen, die andern durch ein mathematisch fundiertes theoretisches Modell. Gerda Ruge, in der freien Wirtschaft tätige Linguistin, sieht den akademischen Methodenstreit gelassen:

Zuspielung (Ruge)

"Eine Tendenz, die in den Staaten sehr stark ist, ist die Corpusanalyse, große Datenmengen herzunehmen und zu versuchen, aus diesen Daten mit statistischen Mitteln etwas herauszubekommen. Daran orientiert sich jetzt auch die Computerlinguistik. Es gibt immer in der Forschung so Strömungen, da glauben alle ein paar Jahre, das muß man jetzt SO machen und gehen alle in DIE Richtung, und ein paar Jahre später kommt wieder was anderes, und nach einer gewissen Zeit, 10 oder 15 Jahre, wird das Alte wieder neu aufgelegt. Man muß da vorsichtig sein, das hat auch was mit Mode zu tun, daß das, was man erreichen wollte, nicht erreicht worden ist und man meint, man muß jetzt anders rangehen. Ich denke, der richtige Weg ist, alles zu tun. Sowohl theoretisch zu arbeiten als auch die Corpora anzuschauen."

Sprecherin
Es wird noch lange dauern, bis der Computer erfunden wird, der die menschliche Sprachfähigkeit in ihrem ganzen Umfang besitzt. Eine Nische aber gibt es heute schon, die eine perfekte Interaktion von Sprache und Computer bietet: Die nonverbale Sprachanalyse mit dem Ziel, den Menschen zum besseren Sprechen zu erziehen. Roland Wagner, Dozent an der Pädagogischen Hochschule Heidelberg, arbeitet seit Jahren erfolgreich mit dem System Audit E.

Zuspielung (Wagner)
"Es geht ja bei uns in der Sprecherziehung darum, daß Personen ihre normale Ausdrucksfähigkeit verbessern können und bestimmte kommunikationsstörende Faktoren abgebaut werden können. Und da haben wir sehr oft mit Leuten zu tun, die nicht besonders gut mit dem Ohr unterscheiden können, was sie jetzt sagen, z.B. Schwerhörige, aber auch Leute, die von einer frappierenden Unmusikalität sind, die nicht mal wissen, ob ein Ton tiefer oder höher ist."

Sprecherin
Wagners Kunden sind junge Lehrer, aber auch Schauspieler und Rundfunksprecher. Es geht um piepsige Stimmen, um Näseln und Nuscheln, um ausländische Akzente und um Dialekte.

Sprecher
Der Vorteil des Computers: Der Benutzer kann seine Äußerungen nicht nur hören, sondern auch sehen. Die Tonsignale werden graphisch und farbig umgesetzt: Die Lautstärke wird ebenso dargestellt wie die einzelnen Frequenzen. Man sieht bei einem bestimmten Laut, wieviel Frequenzanteile im Bereich von 200-300 Hertz vertreten sind und wieviele etwa im Bereich von 8000-10.000. Wer den eigenen Fehler nicht hören kann, weil er eben schwerhörig oder einfach akustisch unsensibel ist, der kann ihn sehen.

Sprecherin
Die junge Schwäbin, die bei der Aufnahmeprüfung zur Schauspielschule durchgefallen ist, weil sie die Ehre des Soldaten genauso aussprach wie die Ähre auf dem Kornfeld, kann an der Amplitude genau erkennen, ob es diesmal richtig war.

Zuspielung (Wagner)
"Es gibt hier einen Fachausdruck, der heißt Formant. Jeder Sprachlaut hat charakteristische Frequenzen, wir sagen Formanten dazu, und diese Formanten lassen sich auf dem Bildschirm sichtbar machen. Ein Student hat im Rahmen seiner Diplomarbeit mal mit diesem System gearbeitet und er konnte recht gut nachweisen, was einen guten Sänger von einem schlechten Sänger unterscheidet. Es sind gewisse Formanten, die gerade in den hohen Tönen den Ausschlag geben, ob eine Stimme als strahlend empfunden wird oder als matt und dumpf. Dann ist man als Sänger nicht mehr nur auf das subjektive Urteil der Lehrkraft angewiesen, sondern man kann am Bildschirm entsprechende Veränderungen bewußt beobachten."

Sprecherin
Das gilt natürlich nicht nur für Sänger, sondern für alle, die ihre Stimme verbessern wollen.

Zuspielung (Wagner)
"Es geht keinesfalls darum, den ausgebildeten Lehrer zu ersetzen, sondern es ist eine zusätzliche Hilfe. Eine Hilfe, die sehr praktisch ist, wenn es um das individuelle, routinemäßige Üben geht. Da kann man alleine zuhause üben und man hat nicht einen unruhig werdenden Lehrer daneben und man hat größere Sicherheit. Das heißt, er ist ein ideales Zusatzmedium für einen unterrichtenden Menschen, kann ihn aber nicht ersetzen."

Sprecherin
Das unterschreiben auch die Computerlinguisten: Die automatische Sprachverarbeitung wird unser Leben erleichtern, aber sicher nicht die Qualität der menschlichen Rede erreichen. Wolfgang Hoeppner zieht das Resumee:

Zuspielung (Hoeppner)
"Zu den Zukunftsperspektiven könnte man sagen, daß eine wichtige Sache die. ist, daß man Sprache nicht isoliert betrachtet, sondern versucht, sie mit anderen Medien, mit Musik, mit Film, mit Bildern in Verbindung zu bringen und damit zum Beispiel Bedienungsanleitungen besser machen kann als sie heute sind. Ich glaube, der Computer kann gar nichts besser tun in einem Bereich, wo Menschen tätig sind."

NUTZUNG VON CL-WERKZEUGEN

Umfrage nach dem industriellen Einsatz computerlinguistischer Software

Wolfgang Hoeppner, Gerhard-Mercator Universität Duisburg
Nils Lenke, Philips-Forschungszentrum, Aachen

Im Herbst vergangenen Jahres hatte die Duisburger Computerlinguistik einen Fragebogen an Industrieunternehmen unserer Region verschickt. Ziel dieser Aktion war es, zu erkunden, inwieweit computerlinguistische Systeme in der Praxis genutzt werden, wie weit sie akzeptiert werden, und u. U. herauszufinden, ob sich eine Kooperation mit industriellen Anwendern ergeben könnte. Das Ergebnis dieser Umfrage hatten wir kurz in unserem Jahresbericht 1994/95 *vorgestellt. Gerhard Knorz bat uns dann, diese Ergebnisse auch für das LDV-Forum zur Verfügung zu stellen. Dem kommen wir gerne mit dem folgenden, leicht überarbeiteten Beitrag nach.*

Es wurden insgesamt 177 Fragebögen verschickt. 3 kamen als Irrläufer, 11 wegen falscher Postadresse zurück. 166 verschickte Fragebögen galten deshalb als Grundgesamtheit. Davon kamen 42 sofort und weitere 29 nach einer "Erinnerung" ausgefüllt zurück, sodaß sich eine Rücklaufquote von immerhin 42,8 % (71 von 166) ergibt.

Angeschrieben wurden schwerpunktmäßig Unternehmen der Industrie, Maschinen- und Anlagenbau sowie die Grundstoffindustrie. Dies spiegelt sich auch in den Antworten:

Branche:

Maschinenbau, Anlagenbau u. a.:	22
Grundstoffe, Montan, Chemie u. a.:	20
Dienstleistung:	6
Entsorgung:	4
Sonstige:	19

Von der Größe her überwiegen die Unternehmen des "Mittelstandes" mit 51-500 Beschäftigten:

Beschäftigte:

1-5:	1
6-50:	13
51-500:	37
500:	20

Von den 71 Unternehmen gaben 8 an, keinerlei Dokumente in einer Fremdsprache zu erstellen und füllten den "Rest" des Fragebogens nicht aus. Die restlichen 63 machten zunächst Angaben über die Art der erstellten Dokumente und die verwendeten Sprachen (Mehrfachnennungen waren möglich):

Textsorten:

Geschäftsbriefe:	56
Rechnungen:	34
Werbeprospekte:	38
Technische Dokumentation:	32

Bei den Sprachen dominiert erwartungsgemäß Englisch, überraschend stark Russisch mit 10 Nennungen:

Sprachen:	
Englisch:	61
Französisch:	32
Spanisch:	13
Russisch:	10
Italienisch:	8
Niederländisch:	4
Ungarisch:	3
Polnisch:	2
Tschechisch:	2
Griechisch:	2
Norwegisch:	2
Portugiesisch:	2
Türkisch:	2
Slowakisch:	1
Schwedisch:	1

Dies kommt auch bei der Angabe von Seitenzahlen zum Ausdruck. Da jedoch nur ein Teil der Unternehmen detaillierte Auskünfte machte, wird auf eine numerische Auswertung verzichtet. Die Angaben schwanken zwischen ,,20" und ,,25000" pro Sprache und Jahr, die höchsten Werte betreffen in der Regel Englisch.

Wer erstellt die Übersetzungen! Dokumente? Antworten zur Auswahl waren: " Fachmitarbeiter, die auch die Fremdsprache beherrschen", "Technische Redakteure und" Übersetzer im Haus" und "Fremde Übersetzungsbüros". Genannt wurde die erste Möglichkeit 51 mal, die zweite nur 20 mal und die dritte 42 mal.

Werden bereits computerunterstützte Verfahren eingesetzt? Das ist nur ganz selten der Fall: 7 Unternehmen antworteten mit "ja". Am häufigsten handelt es sich dabei um Terminologiedatenbanken (3 Nennungen), einmal um ein elektronisches Wörterbuch (Hexaglott EG 4000) und einmal um das Mü-System "METAL" (nämlich beim (früheren) Vertreiber SNI!), zwei Unternehmen machten keine genauen Angaben.

Wie ist man mit der Qualität der Dokumente zufrieden und wie mit den Kosten und der Zeitnähe?

Qualität:	
sehr gut:	13
ziemlich gut:	37
teils, teils: ziemlich schlecht: sehr schlecht:	13
Kosten und Zeitnähe:	
sehr gut:	9
ziemlich gut:	33
teils, teils:	17
ziemlich schlecht: sehr schlecht:	1
Keine Angaben:	3

Tendenziell ist man also mit der Pünktlichkeit und den Kosten weniger zufrieden als mit der Qualität. Ein Versuch, computergestützte Verfahren "an den Mann zu bringen", müßte also hier ansetzen.

Von ~.en 63 Unternehmen haben jedoch nur 11 Änderungen für die Zukunft vor, 2 machten keine Angaben, die restlichen 50 planen keine Veränderung. Jedoch zeigten sich 38 Unternehmen an computerunterstützten Verfahren "interessiert", 22 verneinten dies, 3 machten keine Angaben. Die Frage ist, wieviel dieses "Interesse" wert ist, da auch ein Großteil der interessierten Unternehmen ja keine Änderungen plant.

Unsere Absicht, durch die Umfrage auch Kontakte für eine Kooperation zu gewinnen, ist nicht in dem erhofften Maße erfüllt worden. Dies könnte durch gezielte Ansprache "interessanter" Industriepartner jedoch gelingen, und vielleicht packen wir das auch noch einmal an. Durch den Wechsel von Nils Lenke in das Philips-Forschungszentrum ist eine Motivationsquelle für solche Unternehmungen nicht mehr in Duisburg, und die anderen Mitglieder sind z. Zt. mit zahlreichen anderen Forschungstätigkeiten eingedeckt, daß hier in naher Zukunft wohl keine Aktivitäten entwickelt werden können.

Anschrift:
Gerhard-Mercator Universität Duisburg FB3 - Computerlinguistik
D-47048 Duisburg
Tel: (0203) 379-2006/2008
email: hoeppner@unidui.uni-duisburg.de

Rezensionen

Ray C. **Dougherty:**
Natural Language Computing. An English Generative Grammar in Prolog. Lawrence Erlbaum Associates, Hillsdayle, New Jersey 1994; 349 S. Paperback; ISBN 08058-1526-0

Dieses Buch basiert zum Teil auf Kursen die am Linguistics Department der New York University gehalten wurden. Der Autor versucht eine allgemeine Einführung in die automatische Analyse natürlicher Sprache (Natural Language Processing) mit Prolog zu geben, und legt dabei einen besonderen Schwerpunkt auf die linguistischen Theorien Noam Chomskys.

Die ersten fünf Kapitel bestehen aus einer sehr allgemeinen Einführung in Prolog und die Benutzung von Prolog auf einem Computer. Auch auf die Motivation computerlinguistischer Anwendungen und deren Bezug zu anderen Wissenschaften wird kurz eingegangen. Die restlichen Kapitel sind der Anwendung von Prolog auf linguistische Probleme gewidmet. Kapitel sechs und sieben beschäftigen sich hierbei mit Problemen auf der Wortebene, Kapitel acht mit der Analyse von englischen Sätzen.

Natural Language Computing ist als eine sehr grundlegende Einführung in NLP mit Prolog zu betrachten. Der Autor versucht Lesern ohne jegliche Vorkenntnisse die Grundzüge von Prolog, generativer Grammatik, Linguistik und Computerlinguistik zu vermitteln, kommt dabei aber über eine Darstellung der Grundideen nicht hinaus.

Die dargestellten Programme sind durchgehend sehr einfach und kurz gehalten, und im Verlauf des Buches (welches immerhin 349 Seiten umfaßt) wird kein einziges Programm vorgestellt welches über das Niveau der ersten Kapitel eines Einführungsbuches in Prolog (wie z. B Clocksin und Mellish (1984)) hinausgeht.

Die Einführung in Prolog ist praktisch orientiert, und erklärt ausführlich und detailliert die Funktionsweise dieser Programmiersprache. Wesentliche Aspekte wie Datenstrukturen, Kontrolle des Backtracking, Parsingstrategie oder Unifikation werden jedoch nur unvollständig und unzureichend behandelt.

In Kapitel sechs wird kurz die Lexikalisierung von englischen Verben skizziert. Kapitel sieben behandelt morphologische Phänomene wie Suffigierung und Affigierung. Die im vorigen Kapitel werden dabei nicht verwendet.
Das letzte Kapitel, welches knapp hundert Seiten umfaßt, gibt eine Einführung in die Analyse englischer Sätze. Ärgerlich ist in diesem Zusammenhang die Tatsache, daß im Vorwort die Behauptung aufstellt wird, die dargestellten Programme würden den aktuellen Forschungsstand darstellen (page xiii: 'almost all of the examples encoded into Prolog come from current research in linguistics'). Den einzigen Bezug zu aktuellen linguistischen Theorien stellt die Darstellung des Minimalist Approach (aktuelle linguistische
Theorie Noam Chomskys, (Chomsky 1992)) dar. Diese Theorie wird zwar auf neun Seiten kurz erklärt, die Ideen werden aber nicht ein Programm umgesetzt. Vielmehr nimmt Dougherty die Abschaffung der D-Struktur in Chomsky (1992) als linguistische Begründung für seine kontextfreien Grammatiken in Kapitel acht her. Diese entsprechen formal den in Pereira und Shieber (1987) vorgestellten DCGs (Definite Clause Grammars) in Prolog.

Die dargestellten Grammatiken bestehen aus kontextfreien Regeln ohne Merkmale und sind somit auf dem linguistischen Forschungstand

von 1957. Die in Kapitel 6 und 7 dargestellten Möglichkeiten der Lexikalisierung von Wörtern werden nicht verwendet, und Lexikoneinträge werden nur über ein einziges nichtterminales Symbol beschrieben, so daß z. B. die verschiedenen Subkategorisierungsmöglichkeiten von Verben durch die terminalen Symbole v1 bis v9 dargestellt werden. Der komplexeste Beispielsatz ist hierbei 'the girl gives the truck to the boy'. Die komplexeste vorgestellte Grammatik umfaßt 12 kontextfreie Regeln, 9 Regeln für 9 verschiedene Verbalphrasen (eine für jeden Verbtyp), und jeweils eine Regel für PPs, NPs, und S.

Der Aufbau und die Ausgabe des parse-tree in den Grammatiken ist dabei nicht optimal gelöst (nichtterminale Symbole einer Regel werden mit write innerhalb der Regel ausgegeben). Zudem geht der Autor zwar auf Aspekte wie top-down und bottom-up parsing, leere Kategorien und UnifIkation ein, diese werden jedoch nur oberflächlich erklärt und nicht in Programme umgesetzt.

Insgesamt entsteht der Eindruck, als habe der Autor zwar ein umfangreiches Wissen im Bereich der generativen Grammatikforschung, dafür aber wenig Kenntnisse auf dem Gebiet der Computerlinguistik oder der Programmierung in Prolog. Das dargestellte Material, mag für einen Einführungskurs für Erstsemester ausreichen, das Buch erweckt aber insgesamt den Eindruck als wurden lediglich verschiedene Kursmaterialien verarbeitet. Dies wird dadurch noch verstärkt, daß z.B. in Kapitel 8 nicht auf dem Wissen vorheriger Kapitel aufgebaut wird. Andere Bücher über Prolog, wie z.B. Bratko (1990), gehen auf wenigen Seiten intensiver und mit Hilfe von wesentlich eleganteren Programmen auf NLP ein.

Lobenswert sind allerdings die zahlreichen wunderschönen Schaubilder (ehemalige Overhead-Folien?), die zum Glück auch auf der beiliegenden Diskette (welche alle Programme und einen Shareware Prolog-Interpreter für PC und MAC enthält) als eps-Dateien vorliegen. Als allgemeine Einführung in NLP aber erweckt dieses Buch aber sehr zwiespältige Gefühle. Weder die Prolog-Kapitel noch die dargestellten Programme oder die Einführung in die Theorie der generativen Grammatik können überzeugen. Studenten der Geistes- wissenschaften die keinerlei Vorkenntnisse in Bereich NLP oder im Um- gang mit Computern besitzen mögen vielleicht dennoch das sehr behutsame Heranführen an die Materie zu schätzen wissen.

Literatur

Bratko, J. (1990). PROLOG programming for artificial intelligence. Addison-Wesley.

Chomsky, N. (1992). A minimalist program for linguistic the OIY. Technical report, MIT Worldng Papers in Linguistic.

Clocksin, W. und C. Mellish (1984). Programming in Prolog, 2nd ed. Springer.

Gazdar, G. und C. Mellish (1989). Natural language processing in Prolog. Addison Wesley.

Pereira, F. und S. Shieber (1987). Prolog and Natural-Language Analysis. CSLI.

Marco Zierl, Computerlinguistik Erlangen-Nürnberg

Gerda Ruge:
Wortbedeutung und Termassoziation. Methoden zur automatischen semantischen Klassifikation. Olms Verlag, 1995.

In dem neu erschienenen Buch von Gerda Ruge geht es um ein Verfahren zur vollautomatischen semantischen Klassifikation von Wörtern. Bei dem dargestellten Ansatz ist besonders interessant, daß die für die Bedeutungsähnlichkeit verantwortlichen Relationen aus großen Textkorpora generiert werden. Die aus der Dependenzgrammatik stammenden sog. Head/ModifIer-Relationen drücken Beziehungen zwischen zwei Wörtern aus, von denen eines das andere näher bestimmt. Gemessen wird dann die Übereinstimmung der Heads und Modifiers von je zwei Wörtern. Die sich daraus ergebende Maßzahl kann als Indikator für die semantische Ähnlichkeit zweier Wörter angesehen werden, voraus ge-

setzt die zugrundeliegenden Textkorpora sind groß genug. Dieses ist das Thema des Buches, das sieben Kapitel, einige Anhänge, ein Glossar, ein umfangreiches Literaturverzeichnis und insgesamt 244 Seiten umfaßt. Gleichzeitig beinhaltet dies auch den Neuheitswert, da das Buch als Dissertation an der TU München entstanden ist.

Ich möchte zunächst kurz darstellen, worum es in dem Buch von G. Ruge genauer geht und anschließend darauf eingehen, warum ich glaube, daß dessen Lektüre für Leute unterschiedlichster wissenschaftlicher Herkunft und mit verschiedensten Interessensschwerpunkten einen Gewinn darstellen kann.

Im ersten Kapitel wird zunächst ein kleines Szenario aufgebaut, das jedem vertraut ist, der Literatursuche betreibt. Retrieval-System-Benutzer haben Probleme, sich semantisch äquivalente Schlagwörter einfallen zu lassen. Ein bekanntes Information-Retrieval(IR)-Problem! Wenn hier ein intelligentes System Vorschläge machen kann, ist das sicher sehr nützlich. Damit wird auch die praktische Relevanz des Ansatzes gleich zu Beginn deutlich, wobei die Einbettung in den IR-Kontext eine von mehreren Anwendungsmöglichkeiten darstellt. Im Buch bildet die semantische Klassifikation als IR-Hilfsmittel einen Evaluierungskontext, in dem der dargestellte Ansatz bewertet wird.

Kapitel 2 geht auf die sprachphilosophischen Grundlagen semantischer Ähnlichkeit ein. Es enthält eine kritische Betrachtung des Aspekts Wortbedeutung innerhalb verschiedener Semantiktheorien (der modelltheoretischen Semantik, der Merkmalssemantik und des Wittgensteinschen Bedeutungsbegriffs). Mögliche Vergleichsrelationen werden diskutiert, die zu verschiedenen Auffassungen der Begriffe Synonymie und Ähnlichkeit führen. Die Autorin führt die einzelnen Kriterien der Synonymie- und Ähnlichkeitsdefinition auf, wägt sie gegeneinander ab und entscheidet sich in dem von ihr gewählten pragmatischen Kontext für den Head/Modifier-Ansatz, d.h. auf Synonymie aus der Übereinstimmung aller Heads und Modifiers zweier Wörter zu schließen und den Grad der Ähnlichkeit zweier Wörter vom Übereinstimmungsgrad deren Heads und Modifiers abhängig zu machen. Diese Lösung wird theoretisch aus der Dependenzgrammatik abgeleitet und mit Beispielen belegt.

Kapitel 3 verdeutlicht die Vorgehensweise ausgehend von den Ideen des logischen Empirismus. Es wird ein Axiomensystem entworfen, das nicht die Zusammenhänge zwischen den Bedeutungen einzelner Wörter beschreibt, sondern den Rahmen, innerhalb dessen diese liegen können. Daraus ergibt sich die Möglichkeit, inhaltliche Beziehungen zwischen Relationen zu formulieren und zu formalisieren, ohne die Inhalte im einzelnen zu kennen.

Im vierten Kapitel erhält man einen Überblick über frühere Verfahren zur semantischen Klassifikation, wobei der Bezug zum Neuansatz hergestellt wird. Im wesentlichen stehen dabei korpusbasierte Verfahren sog. lexikonbasierten gegenüber.

Anschließend folgt ein experimenteller Teil, der auf einem Textkorpus amerikanischer Patente basiert, das insgesamt 195.000 Abstracts umfaßt. Hier werden die Experimente beschrieben, welche zum einen die Grundhypothese, die Tragfähigkeit des HeadJ!\.10difier-Ansatzes für die semantische Ähnlichkeitsbestimmung, überprüfen, aber auch das Verhalten von Head/Modifier - Vergleichen in großen Korpora umfassend untersuchen sollen. Enthalten ist hier auch die empirische, methodisch unkonventionelle Bestimmung eines adäquaten Ähnlichkeitsmaßes für Head/ Modifier- Vektoren, welches das menschliche Ähnlichkeitsempfinden widerspiegelt. Aus den Ergebnissen, die belegen, daß mit wachsender Head- bzw. Modisier-Übereinstimmung die semantische Verwandtschaft steigt, leitet sich das Verfahren zur Term-Assoziation als IR-Tool ab, welches das System REALIST mit bereits bestehender morphologischer und syntaktischer Komponente um den Semantikteil ergänzt.

Dieses Kapitel ist m. E. von besonderer Bedeutung für das Buch, da es durch unkonventionellen, aber überzeugenden Versuchsaufbau zeigt, daß ein Ansatz wie der vorgeschlagene nicht nur theoretisch greift, sondern auch in praktischen Umgebungen funktionieren kann.

In Kapitel 6 wird ein Exkurs in psychologische Richtung unternommen. Es wird eine Berechnungsmethode des Head/Modifier-Vergleichs auf der Basis von Spreading-Activation-Netzen vorgeschlagen und der Zusammenhang zwischen automatisch berechneten Assoziationen und dem menschlichen Assoziieren verdeutlicht.

Das Buch schließt mit einem Ausblick, in dem konzeptuelle Hinweise auf weitere wünschenswerte Verbesserungen gegeben werden. Wichtig ist hier die Differenzierung der erzeugten Assoziationen nach Synonymie, Antonymie, Hyponymie etc. Angesprochen werden hier auch Disambiguierung von Polysemen, semantische Relationen zwischen Mehrwortbegriffen und die Behandlung von Abkürzungen.

Wenn ich das Buch von G. Ruge uneingeschränkt zur Lektüre empfehle, so in erster Linie wegen folgender Aspekte:

Der Autorin ist es m.E. gelungen, die Tragfähigkeit des Korpusansatzes im vorgegebenen Kontext nachzuweisen, wobei sie nicht in traditionellen methodischen Grundlagen verharrt, sondern pragmatisch ausgerichtete Konzepte verfolgt, die sehr erfolgversprechend erscheinen. Automatische Wissensextraktion aus unformatiertem Fließtext (im Gegensatz zu strukturierten Lexika) funktioniert, wobei relativ einfache syntaktische Verfahren semantisches Wissen extrahieren können.

Der Ansatz ist trotz seiner überzeugenden theoretischen Fundierung praxisrelevant. Bei dem gewählten Anwendungsgebiet Information Retrieval kann das automatische Klassifikationsverfahren die Funktion eines Thesaurus übernehmen, der den Benutzer eines Retrievalsystems mit alternativen Schlagwörtern zu den ihm bekannten ausstatten kann. Wichtig ist besonders für Praktiker, daß das System dabei als sog. Retrievalhilfe, nicht als selbst entscheidendes Expertensystem konzipiert ist, da die Lösung des Synonymie-Problems durch Vorschlagen von Alternativtermen erzielt wird, aus welchen der Benutzer auswählen kann.

Das Buch richtet sich durch seine interdisziplinäre Ausrichtung an einen heterogenen Adressatenkreis. Der thematisierte Schnittbereich zwischen Informatik und Computerlinguistik vermag sicherlich auch (Kognitions)psychologen, traditionellen Sprachwissenschaftlern und Informationswissenschaftlern viele interessante Erkenntnisse versprechen.

Christa Womser-Hacker, Regensburg

Gebräuchlichkeit und Benutzerfreundlichkeit - SWD, wohin?

Eine Rezension zu: Sacherschließung in Online-Katalogen. Kommission des Deutschen Bibliotheksinstituts für Erschließung und Katalogmanagement. Expertengruppe Online-Kataloge. Berlin 1994. (dbi-materialien. 132.)

Dem umfassenden Nornierungsanspruch für den Wortschatz der Schlagwortnormdatei (SWD) werden 4 Alternativen gegenübergestellt: 1. Reduktion des Wortschatzes auf Begriffe im Sinne von DIN 2330; 2. Verzicht auf die SWD als Mittel der Benutzerführung; 3. Abschied von der Machbarkeit einer Konkordanzklassifikation; 4. Unterscheidbarkeit von Standardsprache zu Fachsprachen.

Wenn im Zeichen der OPAC-Entwicklung und des Datentauschs bisherige Ergebnisse und Methoden der Sacherschließung auf Basis von RSWK/SWD neuerdings als reformbedürfig von den Regelwerksautoren selbst hinterfragt werden 1, dann werden mit dieser Flucht nach vorn kritische Anregungen aus der Fachliteratur aufgenommen2 und zum

1) Sacherschließung in Online-Katalogen. Berlin, insbes. Kap. 4.1-4.5

2) Zum Komplex der RSWK/SWD-Arbeitsmittel vgl. Winfried Gödert zu: Beispielsammlung zu den Regeln für den Schlagwortkatalog. m: Zeitschrift für Bibliothekswesen und Bibliographie 38 (1991) S. 548-598; Winfried Gödert zu: Praxisregeln. m Zeitschrift für Bibliothekswesen und Bibliographie 40 (1993) S. 182-185.

Zu Konzeption und einzelnen Problemen vgl. W. Umstätter, in: ABI-Technik 11 (1991) S. 277-288; Rezension zu Gerhard Halm, in: Knowledge Organisation 21 (1994) S. 47; U. Ribbert, in: Bibliothek 16 (1992) S. 9-25; Urs Bisig, in: MB.NRW 45 (1995), 3, 272-278; I. Prohl, in: Mittei-

eigenen Anliegen gemacht. Generelles und primäres Ziel einer OPAC-gerechten Revision bestehender RSWK/SWD-Regelungen wird in der "klaren Strukturierung, Konsistenz und Einheitlichkeit der Daten" gesehen3, die dem Benutzer einen optimalen Zugang zum OP AC gewährleisten sollen.

Sollen die Reformabsichten der DBI-Kommission und Expertengruppe/n in punkto Sacherschließung tatsächlich vom Willen der Optimierung des Benutzerzugangs geleitet werden, so wären auch solche Überlegungen anzustellen, die im Rahmen einschlägiger Spezialliteratur bisher nur am Rande verfolgt wurden. Die folgenden Ausführungen wollen dazu beitragen und beziehen sich hauptsächlich auf das Produkt der RSWK, auf die SWD und die damit zusammenhängenden Probleme.

Über den Zusammenhang von Normdateien und Regelwerken generell liegen Klärungen aus bibliothekarischer Praxisperspektive vor4, die weithin unbestritten sein dürften. Im Falle der Normdatei SWD wird der Begriff "Normdatei" aber lediglich dahingehend relativiert, daß er "nicht den Eindruck erwecken sollte, als sei ein Normbegriff die einzig richtige Lösung". Natürlich soll die Ansetzung des Normbegriffs von möglichst vielen Anwendern als 'richtig' empfunden werden, es gibt aber genügend Fälle, in denen zwischen mehreren möglichen Lösungen 'formal' zu entscheiden ist.“5 Diese knappen Hinweise auf das (auslegbare) Kriterium der Akzeptanz und auf die Notlösung einer wenigstens "formalen" Entscheidung bei der normierenden Schlagwortvergabe nach RSWK/SWD tragen allerdings der Tatsache nicht Rechnung, daß zur Erstellung von Begriffssystemen und für das Vorgehen bei der Terminologiearbeit offiziell gültige Vorgehensweisen festgelegt sind6, an denen sich auch der Regelungsanspruch der SWD teilweise orientiert, wenn auch nicht konsequent und die einschlägigen Normen mit ihren Regelungen zitierend.

Andererseits wäre es jedoch verfehlt, die SWD als Produkt regelrechter Normanwendung verstehen zu wollen, ist die SWD doch Ergebnis von Schlagwortvergabe auf Basis der RSWK und bezieht sich auf anfallende Bucherwerbungen in wissenschaftlichen Universalbibliotheken. Vom fachgebundenen Anspruch thesaurusförmiger Abbildung eindeutiger Begriffsbeziehungen7 dürfte bei der SWD mit ihrem universellen Wortschatz eigentlich gar nicht die Rede sein, denn "ein universaler Thesaurus ist zwar zugegebenermaßen faszinierend, aber alle bisherigen Versuche dazu müssen als fehlgeschlagen oder nicht vollendet betrachtet werden.“8.

Wird aber angesichts der beschriebenen Schwierigkeiten mangelnder Konsistenz und verbesserungsbedürftiger terminologischer Kontrolle in der SWD9 weiter das Maximalziel einer "klaren Strukturierung, Konsistenz und Einheitlichkeit der Daten" aufrecht erhalten, so läßt dieses Insistieren auf dem einmal gesetzten Zielt so darauf schließen, daß die Regelwerksautoren das in der Sprachforschung bekannte Spannungsfeld "sprachökonomischer Tendenzen" im Gegensatz zu "Sprachredundanz" nicht als Ursache der der beschriebenen Schwierigkeiten erkannt haben11.

lungsblatt. Verband der Bibliotheken des Landes Nordrhein-Westfalen. 1994. S. 169-194

3) Sacherschließung. (Anm. 1) S. 13

4) K. Haller: Kommunikation, Normung und Kataloge. In: Zeitschrift für Bibliothekswesen und Bibliographie 37 (1990) S.403-421

5) Haller (Anm.4) S. 413

6) Insbesondere folgende Normen betreffen den Bereich der SWD: DIN 2330: Begriffe und Benennungen. Allgemeine Grundsätze; DIN 2331: Begriffssysteme und ihre Darstellung; DIN 2332: Benennen international übereinstimmender Begriffe; DIN 2333: Fachwörterbücher. Stufen der Ausarbeitung; DIN 2335: Sprachenzeichen ; DIN 2336: Lexikographische Zeichen für manuell erstellte Fachwörterbücher. Zitiert nach: DIN-Taschenbuch 153. Berlin 1989.

7) vgl. DIN 1463: Erstellung und Weiterentwicklung von Thesauri. Einsprachige Thesauri. 11. 1987. T. 1. Kap. 2.2, *S.2.*

8) Burkart, Margarete: Dokumentationssprachen, Kap. B 5.4.2.1. In: Grundlagen der praktischen Information und Dokumentation. 3.Aufl. Bd 1. München 1990, S. 165.

9) Sacherschließung.(Anm. 1) S. 61-62

10) Sacherschließung. (Anm.) S. 62: Es werden "ergänzende Begriffsdefinitionen sowie zusätzliche assoziative und hierarchische Begriffsrelationen" für notwendig gehalten.

11) Die folgenden Ausführungen basieren auf Hugo Moser: Sprachnorm und Sprachentwicklung. Zur Rolle sprachökonomischer Tendenzen in der heutigen deutschen -Standardsprache, in: Meyers Enzyklopädisches Lexikon. Bd 22, S. 333-337.

Danach sollen im Dienste der Kommunikation sprachökonomische Anstrengungen zur Effizienz, Wirksamkeit und Leistung der Sprache beitragen und zwar im Sinne von

Mag für die bibliothekarische Sacherschließung auch noch so viel Sprachökonomie wünschbar sein, so müßten doch deren Grenzen in der Einsicht zum Ausdruck kommen, daß mit der Normierung des Wortschatzes allein die immer weiter wirksame Sprachredundanz offenbar nicht zu beseitigen ist.

Es mag vielleicht für RSWK/SWD auch schmerzhaft sein, in der Situation schon eingeführter sprachökonomischer Normdateien und Regelwerke inne zu halten und zu fragen: Geht sprachökonomische Normierung an der offensichtlich nicht kleinzukriegenden Sprachredundanz der OPAC-Rechercheure vielleicht vorbei? Und besteht nicht die Situation des OPAC-Rechercheurs gerade aus mehr oder weniger unklar formulierten Problemen, individuellem Vorwissen, unterschiedlichem Schul- und Ausbildungen, verschiedenen biographischen und beruflichen Voraussetzungen - also typischen Momenten redundanten sprachlichen Verhaltens?

Anstatt weiter normierende Maßnahmen der Sacherschließung an Universalbibliotheken vorantreiben zu wollen, wäre es zumindest denkbar, der nicht wegzuleugnenden Sprachredundanz der OPAC-Rechercheure besser Rechnung zu tragen. 4 Möglichkeiten könnten ins Auge gefaßt werden:

1. Reduktion der SWD auf Basis der sprachlichen Unterscheidung von Begriffen und Benennungen;
2. Verzicht auf die SWD als Mittel der Benutzerführung unter Zuhilfenahme der Multimedia-Enzyklopädien auf CD-ROM;
3. Abschied von der Machbarkeit einer Konkordanzklassifikation;
4. Unterscheidbarkeit von Standardsprache in Absetzung zu Fachsprachen.

zu 1.:

Da offenbar weder begriffliche Konsistenz noch optimale Benutzerführung durch die SWD gegeben ist, sollten diese Ansprüche auch nicht weiter aufrecht erhalten bleiben. Von diesem nur auf den 1. Blick negativem Fazit her könnten aber die Bestimmungen der DIN 233012 nutzbar gemacht werden, auf deren Basis der SWD-Wortschatz nach Begriffs arten und -beziehungen (Kap.4) einerseits und nach Benennungen (Kap. 7) andererseits sortierbar wäre. Würden nämlich DIN-gemäß die SWD-Schlagwortkategorien p, k, g als Benennungen bezeichnet und den Regelungen der Formalerschließung zugeführt werden können, dann würden nur die verbleibenden Kategorien s, z, f in den Regelungsbereich der Sacherschließung fallen. Diese Unterscheidung nach Benennungen und Begriffen würde allerdings die gesamte Philosophie von RSWK/SWD in Frage stellen; wenn auch der Vorteil dabei nicht zu übersehen ist, daß mit DIN 2330 ein verläßliches und handhabbares Regelwerk für das Problem der Begriffsbeziehungen, -merkmale und -definitionen schon vorliegt. Auf diese Weise würde der Normungsanspruch für Sachbegriffe jedenfalls verläßlich eingeschränkt werden können.

zu2.:

Andererseits sollten Vorstellungen nicht außer acht gelassen werden, die den Bereich von Sprachnormung generell hinter sich lassen. Wurde schon in der Vergangenheit die Effektivität der intellektuellen Beschlagwortung, auch unter Bezug auf amerikanische Benutzerstudien, in Zweifel gezogen13, eröffnen sich

ausdrucksbezogener Ökonomie, inhaltbezogener/informationsbezogener Ökonomie und Geltungsökonomie, womit Tendenzen zur Rationalisierung, Vereinheitlichung, Präzisierung, der logischen Ordnung, der Sicherung und Wiederholbarkeit in Grammatik und Wortschatz zu verstehen sind.

Der Sprachökonomie entgegen wirken jedoch redundante Tendenzen, die in abstrakter, bildhafter, bewußt-unbewußt ethischer Ausdrucksweise, als Verdeutlichung, inhaltlicher Anreicherung oder auch Entleerung, als Differenzierung und systematisierendes Zusammenwachsen bestehen.

Trotz aller Bemühungen um sprachliche Ökonomie kommt aber auch die Sprachforschung nicht um die Einsicht herum, daß "die Einheit der deutschen Standardsprache, sieht man von der Orthographie ab, nur annähernd gegeben ist. Es gibt regionale und soziale Subnormen. Dazu kommen gruppensprachliche Unterschiede von Vereinigungen aller Art, z.T. in Verbindung mit fachsprachlichen Einflüssen. Eine Tendenz zur Ökonomie und eine solche zur Redundanz stehen also in Spannung zu aller Zeit und in jeder Sprache".

12) DIN 2330: Begriffe und Benennungen. Allgemeine Grundsätze. DIN 2330. März 1979. Zitiert nach: DIN-Taschenbuch. 153. Berlin 1989.

13) Hitzenberger, Ludwig: Intellektuelle Beschlagwortung versus automatische Stichwortvergabe - eine Evaluierungsstudie - , in: ZfBB, Sonderh. zum 71.Deutschen Bibliothekartag 1982, S.159-168; D. Sene u. M. Nagelsmeier-Linke:

zur Zeit neue Möglichkeiten, die der Benutzerführung dienlich gemacht werden könnten. Auch in Deutschland beginnt sich nämlich der Markt den Multimedia- Enzyklopädien auf CD-ROM zu öffnen14. Eine vergleichende Untersuchung dieser neuen Produkte kommt in der abschließenden Einschätzung zu dem Schluß, daß es "prinzipielle Vorbehalte gegenüber der Möglichkeit einer adäquaten elektronischen Repräsentation enzyklopädischen Wissens gibt."15 In Bezug auf Wissenszusammmenhänge und Aneignungsvorgänge erscheint diese Skepsis durchaus angebracht; die Bedeutung enzyklopädischer Produkte als dem Durchschnittswissen angemessenes Zugangsmedium zu maschinell verarbeitetem bibliothekarischem Titelmaterial müßte aber noch untersucht werden. Ein Beispiel, noch dem traditionellen Medium entnommen, möge das wenigstens in groben Umrissen verdeutlichen:

Der Begriff "Bildung"16 ist in der SWD gekennzeichnet durch das alphabetische Hintereinander von Adjektiva- und Kompositabildungen mit dem Stammwort "...bildung...", wobei lediglich die platonische "Paideia" aus dem Rahmen fällt. Würden nun für die OPAC-Recherche zusätzlich zu den bibliographischen Daten einschließlich ihrer Normdateien auch Stich- und Textworte der Multimedia-Enzyklopädien verfügbar sein, dann könnten recherchierbare Titelelemente auch mit den Worten aus den enzyklopädischen Artikelüberschriften und Artikeltexten abgeglichen werden.

Um beim Beispiel "Bildung" zu bleiben: Die Stichworte in Text und Überschrift des Artikels "Bildung" in den Multimedia-Enzyklopädien würden Bezüge auf die antiken, christlich-biblischen, aufklärerischen, klassisch-romantischen und bürgerlichen Traditionen dieses Begriffs mit ihren wichtigsten Personen, Werken bzw. Begriffen, Fakten und Ereignissen auffindbar werden lassen, bezw. mit den entsprechenden bibliographischen Daten im Titelmaterial eines OP AC verknüpfbar sein. Es wäre also absehbar, daß die Mühen bisheriger Begriffsstrukturierung a la SWD sich langfristig durch erweiterte verbale Abgleichsmöglichkeiten erübrigen könnten, dafür aber erweiterte Zugriffe auf andere als alphabetisch legitimierbare Zusammenhänge geregelt werden müßten. Der im neuen enzyklopädischen/lexikalischen Medium mögliche punktuelle Zugriff auf den gesamten sprachlichen Wortschatz in Form von inhaltlich strukturiertem Wissen einer jeweiligen Epoche entspräche ja in geradezu idealer Weise einem allgemeinen Wissens- und Informationsverhalten17, um das sich die OPAC-Kataloge so sehr bemühen.

zu 3.:

Das zurückhaltende Echo18 auf einen Vorschlag der jüngsten Zeit, eine Konkordanzklassifikation zu finden, bezw. unter Ausschluß von Aspekten/Schlüsseln aus trunkierbaren Notationen zu konstruieren19, rührt an das in der Spezialliteratur sattsam bekannte Problem der Unterschiede monohierarchischer, enumerativer und polyhierarchischer Klassifikationssysteme2o. In den jüngst erschienenen Forderungen nach einer OP AC-gerechten und die verbale Sacherschließung ergänzenden Klassifikation21 wird dieser Diskussionsstand aber nicht aufgegriffen, so dass sich dem Leser eher ein Spektrum der Wünschbarkeiten, weniger eines der Machbarkeiten darbietet. Da überdies eine Über

Stichwort oder Schlagwort. In: Bibliothek aktuell 59 (1991), S. 13-15; Ursula Schulz: Einige Forderungen an die Qualität von Normdateien.T.3: Thematischer Zugriff. In: Bibli0-theksdienst 27 (1993) S. 1160-1180.

14) Gödert, Winfried: Multimedia-Enzyklopädien auf CD-ROM. Eine vergleichende Analyse von Allgemeinenzykl0-pädien. Berlin 1994, Deutsches Bibliotheksinstitut. (Informationsmittel für Bibliotheken IFB. Beiheft 1.)

15) Gödert (Anm.14) S. 41.

16) Schlagwortnormdatei (SWD). Ausg. Oktober 1994. Die Deutsche Bibliothek. Leipzig, Frankfurt/Main Berlin 1994. Fiche 008.

17) Schelsky, Helmut: Das Lexikon - Ein Instrument des modernen Bewußtseins, in: Bertelsmann Briefe 47. Sept. 1966. S. 47.

18) B. Lorenz: Notizen zum Stand klassifikatorischer Arbeit. Ein Diskussionsbeitrag. In: Bibliotheksdienst 28 (1994) S. 870-878.

19) Nöther, Ingo: Modell einer Konkordanzklassifikation für Systematische Kataloge, T. l und 2. In: Bibliotheksdienst 28 (1994)S. 15-33undS.175-187.

20) Weishaupt, Karin: Sacherschließung in Bibliotheken und Bibliographien. T. l: Klassifikatorische Sacherschließung. Frankfurt/Main 1985, Kap. 3-5.

21) Sacherschließung. (Anm. 1) S. 34-36.

nahme der "BasisklassifIkation,,22 von großen Schwierigkeiten einzelner Wissenschaftsfächer begleitet ise3, ist die aktuelle Situation einer vereinheitlichenden Klassifikation durch ein Defizit gekennzeichnet, das zur Zeit nicht behebbar zu sein scheine4.

Da in dieser Situation von einer halbwegs funktionierenden Konkordanzklassifikation und von einem allgemein akzeptierten Klassifikationssystem realistischerweise wohl nicht ausgegangen werden kann, sollten lokal klassifizierte Titelaufnahmen im örtlichen, unterschiedlich klassifizierte im regionalen OPAC besser in Kauf genommen werden. Würden nämlich mit gewachsenem Speicherplatz sämtliche vorkommenden Klassifikationssysteme als Normdateien im (lokalen und regionalen) OPAC mitgeführt werden können, dann böte/n unterstützende Klassifikationsnotation/en zwar abweichende Klassenbezeichnungen sowie unterschiedliche Spezifikationsgrade für die Ergebnisse der verbalen Sacherschließung. Das Browsing in den zugehörigen Klassifikationstabellen könnte dem recherchierenden Katalogbenutzer jedoch Aufschlußgeben über den fachlichen Kontext und die Tiefenstruktur (verschieden ausgeprägte Über-, Neben- und Unterordnung der Klassen). Um bei den mitgeführten Klassifikationssystemen eine annähernd genaue Trunkierungsmöglichkeit überhaupt erst zu schaffen, müßten allerdings die hierarchisch ordnenden Notationsbestandteile von den aspektmäßig ordnenden (Anhängezahlen, Schlüssel u. a.) definitorisch getrennt werden - eine Aufgabe, die als Notationsanalyse durchzuführen wäre, dabei auf enumerative Klassiifikationssysteme treffen würde, deren Klassenbezeichnungen gerade durch die unauflösbare Integration von Klassenhierarchie und Anhängezahlen gekennzeichnet sind25

zu 4.:

Die wiederholte Beschwörung der "Gebräuchlichkeit" *(S.* 36,54) und der "Benutzerfreundlichkeit" *(S.* 59) bei einer OPAC-gerechten Schlagwortvergabe wirft, terminologisch gesehen, das Problem der Hoch- oder Standardsprache auf, die als eine "über Mundarten, Umgangssprache und Gruppensprachen stehende allgemeinverbindliche (genormte) Sprachform"26 bezeichnet wird. Zur Festlegung des darunter zu verstehenden Wortschatzes ist daher für die SWD eine Rangfolge primär zu benutzender Nachschlagewerke erarbeitet worden27, in deren Zusammenhang für die Sachbegriffe die Benutzung "Allgemeiner Nachschlagewerke" an 1. Stelle vorgeschrieben ist. Diese Regelung impliziert, daß der Wortschatz der "Allgemeinen Nachschlagewerke" am ehesten (und zurecht) als Garant für die "Gebräuchlichkeit und Benutzerfreundlichkeit" angesehen werden kann. Umso mehr erstaunt es dann aber den Benutzer der SWD, wenn abweichend von den vorkommenden Stichworten in Meyer, Brockhaus und Duden durchaus Bezug genommen wird auf sekundär vorgeschriebene Fachlexika und Fachthesauri29, und wenn Formulierungen der "Allgemeinen Nachschlagewerke" nur als Verweisungsbe-

22) Basisklassifikation für den BibliotheksverbW1d Sachsen-Anhalt. PICA-Projekt Niedersachsen, Projektgruppe F: SacherschließW1g. Erstausg. November 1992.

23) vgl. dazu die problemorientierten Beiträge von R. Oberschelp, E. Bartsch, F. Hillsmann, H.J. Kemchen, H.J. Zerbst, S. Hübner, U. Schulz, R. Baum, E. Bartsch in: mb. Mitteilungsblatt der Bibliotheken in Niedersachsen W1d Sachsen-Anhalt. 84/85,86. 1992.

24) Die "Sachgruppen der Deutschen Nationalbibliographie. Leitfaden zu ihrer Vergabe. Die Deutsche Bibliothek. 1994" spiegeln nicht primär fachsystematische Zusammenhänge wider, sondern bestehen aus aneinandergereihten sch1agwortartigen Begriffen (HSG), die lediglich durch obligatorische Und fakultative Nebensachgruppen (NSG) differenzierbar sind, die aber keine systematischen Über-, Unter- und Nebenordnungen darstellen. Eine stufenweise Trunkierung der Notation zur Unterstützung von. Bedeutungskontexten oder für systematisches Browsing ist mit dieser Klassifikationsstruktur nicht möglich.

25) In der Regensburger Aufstellungssystematik finden sich mehrfache Beispiele dieser Art, nur eines sei hier genannt: BO 7005 ist die Klassenbezeichnung für: Missionsgeschichte nach Kontinenten A - Z. Der hier integrierte Geographie-Schlüssel kann nicht von der bedeutungstragenden Notation BO 7005 separiert werden.

26) Meyers Enzyklopädisches Lexikon. Bd 12. 1978, S. 120.

27) Liste der fachlichen Nachschlagewerke zu den Normdateien (SWD, GKD). Bearb.: Die Deutsche Bibliothek. Frankfurt a. M. 1991, Die Deutsche Bibliothek.

28) Liste. (Anm. 27) S. 68 f. Genannt sind hier unter Abschn. 0: B 1986 (Brockhaus Enzyklopädie), M (Meyers Enzyklopädisches Lexikon), B (Brockhaus Enzyklopädie), Du (Duden "Das große Wörterbuch der deutschen Sprache").

29) Z.B. bei "Bildung" (SWD: Q Böhm - gleiches Stichwort in M,B,Du); bei "BildW1gsdefizit" (SWD: Q Böhm gleiches Stichwort in M); bei "Bildungsfernsehen" (SWD: Q Thes.Päd. - gleiches Stichwort in M).

griffe (BS) eingestuft sind3O. Also doch gelegentlich Bevorzugung des Fachvokabulars, unabhängig von der sich in Enzyklopädien widerspiegelnden Standardsprache? Mag diese Inkonsequenz in den traditionellen Katalogen bisher auch keine besondere Rolle gespielt haben, so wird das Problem doch virulent, wenn der genauere Bezug auf "Gebräuchlichkeit und Benutzerfreundlichkeit" durch die OPAC-Entwicklung jetzt unabweisbarer als früher auf die Tagesordnung bibliothekarischer Regelungen gesetzt wird.

Wäre über den konsequenteren Rückgriff auf die "Allgemeinen Nachschlagewerke" eine mehr oder weniger gegebene Allgemeinverständlichkeit der Sachschlagworte zu erreichen, so würde sich infolge einer derartig veränderten Verschlagwortungspraxis das SWD-Vokabular des "engsten Begriffs" zwangsläufig vermindern. Geht man aber von der Tatsache aus, daß die thesaurusförmigen Instrumentarien der Fach- und Spezialliteratur sich seit eh' und je unabhängig von bibliothekarischen Rege1werken entwickelt und vermehrt haben31, so ist mit dieser außerbibliothekarischen Entwicklung sowieso eine Abgrenzung vom Fachvokabular gegeben. Vorschläge, eine Koordination von bibliothekarischem Schlagwortmaterial mit Wissensbanken und Expertensystemen zu erproben, sind 1985 auf den Widerstand deutscher Bibliothekare gestoßen, sodaß diese Vorstellungen als gescheitert angesehen werden müssen.32 Stände hingegen mit der Nutzung enzyklopädischer Stichworte tatsächlich ein "Gebräuchlichkeit und Benutzerfreundlichkeit" repräsentierender Wortschatz für Katalogbenutzer in wiss. Allgemeinbibliotheken bereit, würden sich die in der jüngsten Veröffentlichung beschworenen Normierungsanspruche für die Sacherschließung als Benutzerzugang sicherlich nicht mehr aufrechterhalten lassen. Für die Verschlagwortung aktueller und innovativer Sachverhalte in der publizierten und katalogisierten Literatur, die über den Wortschatz der "Allgemeinen Nachschlagewerke" hinausgeht, griff auch die SWD nicht selten zur vorliegenden, analogen oder autorisierten Ansetzungsformen33, was in Kombination mit den übrigen für die OPAC-Recherche zur Verfügung stehenden Titelelementen der Formalerschließung sicherlich eine plausible Vorgehensweise ist. Die wichtige Rolle einer unterstützenden klassifikatorischen Sacherschließung im Falle innovativer und aktueller Literatur muß an dieser Stelle nicht extra hervorgehoben werden, auch wenn kein einheitliches bezw. konkordierendes Klassifikationssystem in Sicht ist.

Die erstmals mögliche Konfrontation von Sachbegriffen aus katalogisierten Publikationen mit dem Wortschatz der das Durchschnittsverständnis repräsentierenden Multimedia-Enzyklopädien wurde in der bibliothekarischen Fachdiskussion bisher noch nicht diskutiert; die einleuchtend beschriebenen Schwierigkeiten weitergehender Normierung im Bereich der Sacherschließung legen es aber nahe, die hier skizzierten Anregungen wenigstens für die inhaltliche Erschließung über Sachbegriffe nach RSWK/SWD zum Beratungsgegenstand des DBI zu machen.

Gisela Hartwieg, Berlin

30) Z.B. bei "Lehrmittel" (SWD: Q Böhm "Bildungsmittel" in M); bei "Unterrichtstechnologie" (SWD: Q M, W Päd - "Bildungstechnologie" neben "UnterrichtstechnoIogie" in M, *also* keine Synonyme wie lt SWD); bei "Erziehungsziel" (SWD: Q SWL - "Bildungsziel" in M). *Auf* die Möglichkeit, den Wortschatz der SWD auf Basis von Enzyklopädien/Lexika um relevante Begriffe zu ergänzen, soll nur mit wenigen Beispielen aufmerksam gemacht werden. So fehlen in der SWD z.B. Bildungsbegriff (M), Bildungsdichtung (M), bildungsfähig (Wahrig u. D), bildungsfeindlich (D), Bildungsgefälle (M), Bildungsgehalte (M) u. ä. m.

31) Einen Überblick bietet *"Who* is who in der Online-Szene. Das Jahrbuch der Online-Szene. b team B. Breidenstein GmbH. 1994/5, S. 331-336.

32) Umstätter, Walter: Wäre es nicht langsam Zeit..., in: ABI-Technik. 11. 1991/4, S. 277-288.

33) Allein bei den Einträgen "Bildung.." in der SWD wird als Quelle zitiert: Analogiebildung, analog (4 x), Vorlage (7 x), Autor, Autor DB (12) von insgesamt 83 Quellenangaben, was etwa 25% ausmacht.

2. GLDV-HERBSTSCHULE MODERNE METHODEN DER CORPUSANALYSE 11.-15. SEPTEMBER 1995 IN BONN

Carsten Bredanger

1 GLDV -Herbstschule '95 in Bonn

Zum zweiten Male traf im Frühherbst 1995 eine kleine Gruppe Studierende, DoktorandInnen, Lehrende und Interessierte aus der beruflichen Praxis zusammen, um sich einiger ausgewählter Spezialthemen der Corpusanalyse anzunehmen. Die extrem günstige quantitative Relation zwischen Lehrenden und Lernenden erlaubte die ebenso differenzierte wie unmittelbare Ansprache, Motivation und Herausforderung der individuellen Fähigkeiten und Lernfreude der einzelnen Studierenden durch anschaulich und diskussionsfreudig dargebotenes Lehr- und Übungsmaterial auf hohem fachlichen und didaktischen Niveau. Ebenso trug die konsequente Begrenzung kombinierbarer Seminare auf maximal drei Kurse von je 10 Stunden Umfang, freie Übungszeiten nicht mitgerechnet, zu günstigen Studierverhältnissen bei. Das als Tagungsort gewählte Psychologische Institut der Universität Bonn entsprach hinsichtlich der Ausstattung von Seminar- und Rechnerräumen den Anforderungen der Herbstschule.

2 Moderne Methoden der Corpusanalyse

Sechs Kurse standen den TeilnehmerInnen zur Wahl. Organisiert in drei Kursblöcken, erlaubte das Spektrum angebotener Kurse eine individuelle Schwerpunktsetzung: Roland Hausser (Erlangen- Nürnberg) ließ es sich nicht nehmen, mit "Morphologie und Tagging" ein klassisches computerlinguistisches Thema aufzugreifen und die automatische Wortformenerkennung mit seinem Tagger „LA-MORPH" auch in der Praxis zu demonstrieren.

Karin Haenelt (Darmstadt) profitierte in ihrem Kurs "Textmodellbasierte Korpusanalyse" von ihrer langjährigen Projektleitertätigkeit bei der GMD und zeigte Probleme der Corpusanalyse und lexikalisierter Textgrammatik anhand des KONTEXT-Modells und des Context Feature Structure Systems auf. Als eine Einführung in statistische Auswertungsmethoden über Korpora und in die linguistische Relevanz statistischer Verfahren definierten Robert Neumann, Cyril Belica und Doris al-Wadi (allesamt Mannheim) ihren treffend benannten Kursus "Statistischer Zugriff auf Korpora: Disambiguierung und Tagging". Das Erstellen von Korpora behandelten Henning Bergenholtz (Aarhus, Dänemark) und Randall L. Jones (Provo, USA). Während sich Jones auf "Korpora gesprochener Sprache" spezialisiert hat und die TeilnehmerInnen seines Kurses kenntnisreich in die Thematik der Korpuserstellung einführte, die ihm aus der täglichen Arbeit mit seinem BYU-Corpus nur allzu gut vertraut ist, betonte Bergenholtz Besonderheiten fachsprachlicher Korpora. In diesem Rahmen wurde auch die unter Linguisten umstrittene Trennung zwischen Fach- und Gemeinsprache behandelt. Die sicher nicht einfache Aufgabe, einen Einblick in das komplexe Normenwerk der Standard Generalized Markup Language (SGML) zu geben, war das erklärte Ziel von Peter Scherber (Göttingen) mit seinem Kurs

"Methoden der Standardisierung - Eine Einführung in SGML". Auch die vor allem in linguistischen und literaturwissenschaftlichen Kontexten verwendete TEI (Text Encoding Initiative) wurde, dem Rahmen des Kurses angemessen, in Grundzügen besprochen und anhand von praktischen Beispielen anschaulich dargestellt.

3 Rahmenprogramm

3.1 Eingeladene Vorträge

Von den Organisatoren ins Rahmenprogramm gepackt, aber sicher nicht nur am Rande zu erwähnen sind die beiden im Rahmen der Herbstschule als Plenarvorträge gedachten öffentlichen Abendvorträge. Leider mußte der bereits im Vorfeld mit Spannung erwartete Vortrag von Manfred Bierwisch (Berlin) mit dem programmatischen Titel "Linguistik als Geistes- und Naturwissenschaft" aufgrund einer kurzfristigen Erkrankung zum allgemeinen Bedauern entfallen. Der Positionsvortrag "Wortvernetzungen in Computer und Gehirn" von Helmut Schnelle (Bochum) zeigte in einer detaillierten Analyse strukturelle Aspekte von Sprachverarbeitung und Repräsentation von Sprache in Gehirn und Computer vor dem Hintergrund aktueller neurophysiologischer Erkenntnisse auf. Die von Schnelle postulierte Notwendigkeit der Wende des Sprachbegriffs und der Reinterpretation linguistischer Prozesse als Datenverarbeitungsprozesse unter Zuhilfenahme des erweiterten Datenbegriffs aus der Kybernetik blieb in der Diskussion nicht unkritisiert.

3.2 Exkursion in das GMD-Forschungszentrum Schloß Birlinghoven

Computer und neue Medien verkörpern wertvolle Techniken, dienen oft aber zu wenig mehr als modischen Status- und Trendsymbolen. Um interessierten TeilnehmerInnen der Herbstschule Sinn und Nutzen neuer Medien aufzuzeigen, lud die GMD nach Schloß Birlinghoven. Die Präsentationen von Klaas Sikkel über *Kooperationsunterstützung* in heterogenen Arbeitsumgebungen via Internet (BSCW) und von Karin Haenelt, die die ihrem Herbstschulkurs zugrundeliegende KONTEXT-Implementierung vorführte, stießen auf lebhaftes Interesse. Laborvisiten in den Bereichen "Interaktives Fernsehen" (Ralf Wegner) und "Telekonferenz via Satellit" (Peter Kanzow) ließen die neuen Möglichkeiten anklingen, die sich durch die Nutzung neuer Medien eröffnen. Trotz kleinerer technischer Pannen trübte einzig der eng gesteckte Zeitplan das positive Gesamtbild. Abgerundet wurde der Tag durch eine Wanderung im Siebengebirge mit anschließendem geselligen Weinabend.

4 Fazit

Wie den zufriedenen Mienen der Herbstschul-TeilnehmerInnen am Abschlußtag zu entnehmen war, kann die GLDV mit ihrer Veranstaltung einen Erfolg für sich verbuchen. Nicht zuletzt dazu beigetragen hat sicherlich das hervorragende Preis-/ Leistungsverhältnis der Veranstaltung, daß durch den Verzieht der Herbstschul-Dozentlnnen auf leistungsgerechte Bezahlung ermöglicht wurde.

Es bleibt zu wünschen, daß der erfolgversprechende Weg der Herbstschulen auch weiterhin beschritten wird.

INFORMATIONSMANAGEMENT INFORMATIONSMÄRKTE INFORMATIONSGESELLSCHAFT

Tagung an der Fachhochschule Darmstadt - 10 Jahre Fachbereich Information und Dokumentation

Gerhard Knorz

„10 Jahre Hochschulausbildung für einen Schlüsselbereich wirtschaftlicher Entwicklung" feierte der Fachbereich Information und Dokumentation **mit einer Tagung vom 4. bis 6. Oktober 1995 an der Fachhochschule Darmstadt. Der in Hessen konkurrenzlose informationswissenschaftliche Fachbereich nutzte die Gunst der Stunde, im Kontext der aktuellen öffentlichen Diskussion um den Weg in die Informationsgesellschaft ein fachlich hochkarätiges Vortragsprogramm zusammenzustellen. Eine lichtdurchflutete Industriehalle verlieh der Veranstaltung ein eigenes technisches Ambiente, das nicht repräsentative Sattheit, sondern Aufbruchstimmung und Beweglichkeit signalisierte. Eine Ausstellung mit innovativen Projektarbeiten des Fachbereichs, einer Auswahl von Diplomarbeiten und einem Stand mit EG-Informationen begleitete die Veranstaltung.**

Strategische und politische Aspekte des Informationsbereichs

Die Veranstaltung wurde durch den Rektor der Fachhochschule, *Prof. Dr. Kremer* und den Dekan des Fachbereichs, *Prof. Dr. Knorz*, den Verantwortlichen für Programm und Organisation, eröffnet. Dieser erste Nachmittag war den strategischen und politischen Aspekten des Informationsbereichs gewidmet. Über *"die Rolle der Hochschulen in der Informationsgesellschaft* « sprach Staatssekretär Rolf Praml, der versuchte, einem euphorisch-technokratischen Standpunkt in der gegenwärtigen Diskussion um Datenautobahn und Informationsgesellschaft eine eher distanziert-kritische Auffassung entgegenzusetzen. Durchaus ein Kontrast zur Botschaft von Erika Mann, Mitglied des Europäischen Parlamentes, die engagiert die zentrale und strategische Rolle herausstellte, die der Entwicklung des Informationssektors von der EU zuerkannt wird - als grundlegende Voraussetzung für den Erhalt von Wettbewerbsfähigkeit gegenüber alten und neuen Konkurrenten auf dem Weltmarkt. Nachhaltig warb sie für eine europäische Dimension des Denkens und für eine Beteiligung an den verschiedenen europäischen Förderprogrammen.

Wie schnell sich Einschätzungen wandeln und Innovation in Technik und Strategien die Handelnden in der Informationswirtschaft vorantreibt, während es andererseits noch viele Hindernisse zu überwinden gilt, vermittelte *Arnoud de Kemp* als Vorsitzender der Deutschen Gesellschaft für Dokumentation (DGD) und Vertreter des wissenschaftlichen Springer-Verlags Heidelberg. Speziell berichtete er über das *Medoc-Projekt* einer virtuellen weltweiten Informatik-Bibliothek im Internet, an der Springer, FIZ Karlsruhe und viele andere Partner arbeiten (medoc@informatik.tu-muenchen.de).

Unter dem Titel *„Information Retrie-*

val im Umbruch - belegte *Prof. Dr. Gerhard Knorz* seine Überzeugung, daß die methodische Stagnation der letzten 3 Jahrzehnte, was die Entwicklung von Retrievalwerkzeugen betrifft, ihr Ende erreicht hat. Nicht zuletzt ist die amerikanische Initiative TREC (Text Retrieval Conference), ein offener Wettbewerb zwischen Retrievalverfahren aus Hochschule und Industrie, ein wichtiger Schritt dahin, dass Ranking-basierte Retrievalsysteme die akademische Spielwiese in Richtung Praxis verlassen werden.

In der Diskussion um den tiefgreifenden Wandel im Informationsbereich wird Osteuropa hierzulande weitgehend vergessen. Am Beispiel der OPAC-Entwicklung für die durch Brand zerstörte Universitätsbibliothek Bukarest berichteten Gerhard Riesthuis (Univ. Amsterdam) und Victoria Francu (Universitätsbibliothek Bukarest) über die Erfahrungen und Ergebnisse einer erfolgreichen Zusammenarbeit im Rahmen von TEMPUS und belegten, daß - trotz aller Schwierigkeiten - eine Bibliothek in Rumänien weiter sein kann als so manche große Bibliothek in der Bundesrepublik.

Weites Spektrum von Themen für Informationsspezialisten

Der Donnerstag offerierte eine große Bandbreite von Themen und deckte damit das gesamte Spektrum ab, für das Darmstädter Diplom - Informationswirte/wirtinnen ausgebildet werden: Wirtschaftsinformationen aus Sicht eines Anbieters *(Walter Niedermeyr,* Genios Wirtschaftsdatenbanken), Medieninformation aus der Perspektive eines Newcomers im Pressebereich *(Jutta Oidimann,* FOCUS), Chemie-Information aus Sicht der Chemischen Industrie *(Dr. Gabriele Kirch- Verfuß,* Hüls AG). Es ging um Informationsmanagement, wie es in Beratungs- und "Antrags-"intensiven kleinen Unternehmen unverzichtbar ist *(Folkmer Rasch,* Rasch und Partner, Darmstadt) und wie es in großen Unternehmen in der verteilten Produktion sich gegenwärtig wandelt *(Prof. Dr. Bernd Hamacher,* BIBA Bremen und seit Oktober '95 Professor für Informationsmanagement an der FH Darmstadt). Es ging um Electronic Publishing, speziell um CD-ROM-Produkte, als Herausforderung und Chance für Repress und Press- Unternehmen (Markus Rabsch, Druckhaus Waiblingen) und um die Kommerzialisierung des Internet, speziell um die heute schon realisierbare Gestaltung von interaktiven 3D-Welten für den Konsumenten von morgen (Bauer, ZGDV). Es ging um das (elektronische) Management von Informationen, die papiergebunden schon wegen ihres unglaublichen Umfangs konventionell nicht mehr handhabbar sind *(Anion Schmauiz* sowie Markus Heiß und Dirk Rietzschel, Zusatzversorgungskasse des Baugewerbes) und um das Problem, das Know-How eines weltweit tätigen Beratungsunternehmens so zu managen, daß der Berater in New York ohne langwierige Such- und Kommunikationsprozesse die Erfahrung seiner Kollegen in Frankfurt und Barcelona in sein aktuell zu erstellendes Konzept oder Angebot einbinden kann *(Dr. Jürgen Winkelmann,* A.T. Kearney) .

Die interessierten Teilnehmer konnten insgesamt mitnehmen, daß Internet von niemandem aus der Betrachtung ausgeklammert wird, daß aber dessen gegenwärtige und zukünftige Rolle von Anbietern, Entwicklern und den Nutzern in den Unternehmensberatungen, den Medien und der Industrie recht unterschiedlich ein geschätzt wird. Daß das Outsourcing Konzept der FOCUS-Dokumentation offensichtlich funktioniert, aber ganz entscheidend und unverzichtbar auf die Nutzung eines etablierten Pressearchivs setzt. Daß Management auch und gerade von Information immer mehr als eine Aufgabe begriffen wird, die mit Kommunikation, Abstimmung und Konsensbildung zu tun hat. Daß als Reaktion auf Kosten- und Konkurrenzdruck nicht nur der Abbau von Personal im "nicht-produktiven" Bereich bleibt, sondern daß gerade ein überzeugendes und realisiertes - Informationskonzept mit den entsprechenden Fachleuten entscheidend zum Geschäftserfolg beitragen kann.

Man kann - zumal nach 17:00 Uhr über europäisches Informationsrecht sicher in einer Weise sprechen, die den durch

schnittlichen Interesssierten aus dem Saal treibt. Das Thema zum Abschluß des Tages in einer spannenden und nachvollziehbaren Weise behandelt zu haben ist das Verdienst von *Dr. Jürgen Goebel.*

Bedingungen und Ziele von informationswissenschaftlicher Hochschulausbildung

An dem letzen Freitag Vormittag wollte der Fachbereich Information und Dokumentation das Programm auf seine ureigenen Belange ausrichten. Zunächst stellte *Prof. Dr. Joachim Kind* den aktuellen Diskussionsstand zu einer neuen Strukturierung der informationswissenschaftlichen Ausbildung in Darmstadt vor, die sich bewußt nicht an den geläufigen Etiketten für Lehrinhalte orientiert. Weit über den Hochschulbereich hinaus zielten die Überlegungen von *Ursula Piccolo* (Universität Göttingen), einer Mitautorin der von allen einschlägigen Gesellschaften herausgegebenen Broschüre *"Informationskultur für die Informationsgesellschaft u.* Sie präzisierte, was sie unter einer *"Bildung für die Informationsgesellschaft* versteht und welche Forderungen an Gesellschaft und Politik sich daraus ergeben.

Um die *europäische Dimension von informationswissenschaftlicher Ausbildung* ging es in zwei Beiträgen: *Prof. Dr. Thomas Seeger* moderierte eine lebendige Podiumsdiskussion mit StudentInnen des Fachbereichs, die ihr berufspraktisches Semester in Moskau, Brighton und Luxemburg mit ganz unterschiedlichen, aber durchweg positiven Erfahrungen absolviert hatten. Kein Zweifel, daß Erfahrungen im Ausland viel selbstverständlicher werden sollten. Gerade das ist das Ziel europäischer Förderprogramme wie Sokrates und Leonardo, über die Mitarbeiter aus den Referaten Technologietransfer und Auslandsbeziehungen informierten.

Eine gewisse und zudem positive Überraschung war das Interesse, das der Abschlußdiskussion *"Informationswissenschaft-Quo vadis"* auch über die reguläre Veranstaltungszeit hinaus entgegengebracht wurde. Nach einem einführenden Referat von *Prof. Dr. Rainer Kuhlen* (Universität Konstanz, Informationswissenschaft) diskutierten *Dr. Ilse Harms* (Univ. des Saarlandes, Informationswissenschaft), *Prof. Dr. Jürgen Krause* (Univ. Koblenz, Direktor des IZ Sozialwissenschaften und Vorsitzender des Hochschulverbandes für Informationswissenschaft), *Prof. Dr. Gerhard Knorz, Marlies Ockenfeld* (GMD, IPSI) und *Prof. Dr. Thomas Seeger*. Die Entwicklung des Faches Informatik in den letzten 20 Jahren und speziell das der Wirtschaftsinformation lieferten in Kontrast zu der der Informationswissenschaft genug "Provokation", um Defizite, aber auch Entwicklungssperspektiven von Wissenschaft und Praxis des Informationssektors teils kontrovers aber im Ergebnis keinesfalls pessimistisch zu analysieren.

Resumee

Wenn die Veranstaltung - wie der Dekan einleitend sagte - das Ziel hatte, eine gefällige Selbstdarstellung des Fachbereichs anläßlich seines 10-jährigen Bestehens zu vermeiden um statt dessen die gegenwärtig aktuellen Themen in den Fachbereich hineinzutragen, dann wird man die Tagung als rundum gelungen bezeichnen - mit einem Wermutstropfen: Die studentische Beteiligung war weitaus geringer, als man sie von einem immer noch jungen, innovativen Fachbereich erwarten würde, dessen Ausbildungsziel doch gerade professionelle Neugier ist! Daß studentisches Engagement dennoch kein Fremdwort ist, bewies der Tagungsraum, den Mitarbeiter und Studierende in Eigenarbeit aus einer wartenden Baustelle heraus entwickelt hatten sowie die reibungslose Betreuung der Tagung und Exponate (CD-ROM-Eigenentwicklungen, WWW-Informationsserver, Datenbanken, Informationen zu den europäischen Informationsprogrammen).

Als begleitende Unterlage zu Tagung erhielten die Teilnehmer die vom Gründungsdekan Thomas Seeger herausgegebene und 421 Seiten umfassende Festschrift *"Aspekte der Professionalisierung des Berufsfeldes Information u,* die als Band 21 der HI-Reihe *Schriften zur Informationswissenschaft* im Universitätsverlag Konstanz erschienen ist.

Workshop on Discourse Markers

Duisburg 6./7. Oktober

Prof. Dr. Wolfgang Hoeppner
Gerhard–Mercator–Universität Duisburg

Teilnehmerliste:

Degand, Liesbeth (Louvain, B)
Foolen, Ad (Nijmegen, NL)
Hoeppner, Wolfgang (Duisburg)
Knott, Alistair (Edinburgh, UK)
Lenke, Nils (Duisburg)
Risselada, Rodie (Amsterdam, NL)
Sanders, Ted (Utrecht, NL)
Schmitz, Birte (Berlin)
Spooren, Wilbert (Tilburg, NL)
Stede, Manfred (Berlin / Ulm)

Einige Teilnehmer wurden aus Mitteln des Beauftragten für die Zusammenarbeit von Hochschulen in B, D und NL unterstützt.

Abschlußbericht:

Der Workshop fand am Freitag, dem 6.10. von 12.00 Uhr bis 18.30 Uhr und am Samstag, 7.10., von 9.00 bis 15.00 statt.

Die Teilnehmerzahl des Workshops war absichtlich begrenzt worden, um eine fruchtbare Diskussion zu ermöglichen. Dieses Konzept erwies sich als sehr erfolgreich. Da alle Teilnehmer bereits auf dem Gebiet der Diskursmarker durch Veröffentlichungen hervorgetreten sind und hier jeweils ihre aktuellen Forschungsvorhaben vorstellten, ergaben sich intensive Diskussionen.

Ein zentraler Punkt der Diskussion war die Rolle, die Diskursmarker bei der Aufdeckung verborgener Mechanismen der Text und Diskursproduktion bzw. -rezeption spielen können. Dies gilt sowohl für konnektive Marker, also Konjunktionen usw., die Aufschluß über zugrundeliegende rhetorische Relationen geben können, als auch für Partikeln, die zur Diskursteuerung eingesetzt werden.

Im Einzelnen wurden folgende Vorträge gehalten:

- L. Degand über „Markers for Causality"
- W. Spooren über „The Use of Contrastive Cause-Consequence Relations by Primary School Children"
- R. Risselada über „Adverb And/or Discourse Marker? Latin Nunc Compared to English Now"
- B. Schmitz über „An Inventory of Discourse Functions for Dialogue Interpreting"
- N. Lenke über „Ja, Doch and Denn: Marking Common Knowledge in Dialogues"
- Alistair Knott über „A Set of Orthogonal Features for Describing the Space of Sentence / Clause Connectives"
- Ted Sanders über „Integrating Intentions and Relations"
- Ad Foolen über „Concessive and Adversative"
- M. Stede über „Generating English and German Markers for Concession".

Ein Nachfolgeworkshop ist für 1996 oder 1997 geplant, vermutlich in den Niederlanden.

BERICHT ÜBER DIE TEI-WORKSHOPS AM ZENTRUM FÜR DATENVERARTBEITUNG DER UNIVERSITÄT TÜBINGEN

Winfried Bader

Vom 15.-17. November fand am ZDV der Universität Tübingen ein Workshop zum Thema *SGML-konformen Textauszeichnung nach den Richtlinien der Text Encoding Initiative (TE/)* statt, der infolge des großen Andrangs vom 11.-13. März 1996 wiederholt wurde. Insgesamt 80 Teilnehmer/innen aus ganz Europa, die elektronische Texte einsetzen für kritische Edition von Einzeltexten und Textsammlungen, linguistische Analyse, erzählanalytische Textbeschreibung, Wörterbücher und Lexika, Bibliothekswesen, Sammlung von Einzelartikeln, Bereitstellung von Lehrmitteln, besuchten die beiden Veranstaltungen.

Sie wurde gehalten von Michael Sperberg-McQueen PhD (Chicago, IL, USA), der als leitender Herausgeber die TEI-Richtlinien maßgeblich mitgestaltet hat, und Dr. Winfried Bader (ZDV), Mitglied des Advisory Boards der TEL Das Thema SGML erweckt erst seit der weiten Verbreitung von HTML im WWW allgemeines Interesse; die ISO-Norm selbst existiert aber bereits seit 1986. Auch die Problemstellung, SGML für den Austausch und die elektronische Publikation geisteswissenschaftlicher Texte und ihrer wissenschaftlichen Analyse zu verwenden, beginnt jetzt bei einer größeren Öffentlichkeit Interesse zu finden.

Der folgende Bericht über die Themen der Workshops gibt gleichzeitig einen ersten Überblick über die wichtigsten Ziele und Eigenschaften der TEI:

1 Dokumentenanalyse

Bei diesem Schritt ging es darum, die Phasen der Arbeit mit elektronischen Texten zu zeigen. Wichtig ist, vor jeglicher elektronischer Arbeit sich klar zu machen, daß Textauszeichnung heißt, eine oder mehrere Interpretationen eines Textes explizit zu machen. Die Dokumentenanalyse klärt die inhaltlichen und wissenschaftlichen Ziele, aber auch die Rahmenbedingungen (Geld, Zeit, Ziel), für die Erfassung elektronischer Texte. Als Beispiele dienten eine Akkadische Rationenliste, Autographen eines modernen Autors, ein- und mehrsprachige Wörterbücher, Enzyklopädien, Gedichtanalysen.

2 Einführung in SGML

Die ISO-Norm 8879:1986 Standard Generalized Markup Language ist eine Metasprache für Textauszeichnungssprachen. Ihre Syntax besteht aus Element-, Attribut- und Entity-Vereinbarungen, die konstitutive Bestandteile einer Document Type Definition (DTD) sind. In einer Demonstration wurde das Erstellen einer DTD und eines dazugehörenden Dokumentes gezeigt, das anschließend mit dem Parser sgmls auf SGML-Konformität überprüft wurde.

3 Einführung in TEI

Die Text Encoding Initiative hat Richtlinien entwickelt, in der die Problematik der Auszeichnung geisteswissenschaftlicher Texte aus allen Fachrichtungen berücksichtigt ist, und die gleichzeitig SGML-konforme DTDs bereitstellt, die in jeder SGML-Software verwendet werden können. Eine Übersicht über das Grundinventar der TEI-Auszeichnungen

wurde gegeben und die modulare Struktur der TEI-DTDs samt ihrer SGML-technischen Details erklärt. Auf einer Kerngruppe von Auszeichnungen, die bei allen Textsorten gebraucht werden, bauen Basisgruppen auf, die jeweils für einen bestimmten Texttyp (Prosa, Poesie, Drama, Wörterbuch, ...) Auszeichnungen bereitstellen. Frei kombinierbar kommen die Zusatzgruppen hinzu, die jeweils einer bestimmten Art der Textuntersuchung (Verkettung und Verweise, Analyse, Textkritik, ...) dienen. Mit dem Author/Editor, der dank der Erlaubnis der Firma SoftQuad auch für eigenes Üben im PC-Pool zur Verfügung stand, wurde der praktische Einsatz der TEI-DTD demonstriert.

4 TEI-Auszeichnungen für "Verkettung und Parallelisierung"

Diese Zusatzgruppe der TEI-Auszeichnungen stellt zahlreiche Mechanismen für Querverweise, zur Verkettung von Textstellen, für die Verweismechanismen zum Anbinden von Analysen, zur Parallelisierung verschiedener Dokumente und für Hyperlinks bereit. Sie sind flexibler und umfassender als die Verweismechanismen von HTML (Hyper Text Markup Language).

5 TEI-Auszeichnungen für "Einfache Analyse Mechanismen"

Die Auszeichnungen dieser Zusatzgruppe dienen vor allem Analysen aus dem linguistischen Bereich. Außerdem bietet diese Gruppe interessante Möglichkeiten, Einzelanalysen zu bestimmten Textstellen über Verweismechanismen an getrennter Stelle im Dokument zu verwalten, wodurch eine übersichtlichere Anordnung ermöglicht wird. Es erlaubt auch, schon vorhandene Dokumente nachträglich mit Analysen zu versehen, ohne das Dokument selbst verändern zu müssen.

6 SGML-Software in der Praxis

a) Editoren:
SGML-Editoren sind in der Lage, eine beliebige DTD zu lesen, um dann bei der Erstellung des Dokumentes kontextsensitiv nur die Auswahl der an dieser Stelle entsprechend der DTD legalen Auszeichnungen zu ermöglichen, und sie bequem per Mausklick einzutragen. Demonstriert wurde psgml, ein Zusatz zum emacs (unter LlNUX), Author/Editor und RulesBuilder (zum Kompilieren von DTDs) von SoftQuad (beide unter WINDOWS), und TUSTEP (unter AIX).

b) Parser:
Parser prüfen, ob ein bestimmtes Dokument zu einer gegebenen DTD konform ist. Vorgestellt wurde sgmls (unter LlNUX), der neben der Prüfung als zusätzlichen Output eine Umformung des Dokumentes bietet, die die Weiterverarbeitung erleichtert, die Parsing-Funktion von Author/Editor und eine TUSTEP-Kommandofolge zum Prüfen von HTML-Dokumenten.

c) Browser:
Browser dienen dazu, SGML-Dokumente am Bildschirm in ansprechender Form anzuzeigen, das Blättern innerhalb des Dokumentes zu erleichtern, das Verfolgen von Verweisen (Hyperlinks) sowie eine Volltextsuche zu ermöglichen. Vorgestellt wurde Panorama von SoftQuad, dem allerdings die Suchmöglichkeiten fehlen, DynaText und DynaWeb von EBT. Letzteres bietet die Möglichkeit, beliebige SGML-Dokumente im WWW mit HTML zu präsentieren, trotzdem aber im Hintergrund die volle SGML-Struktur zu haben, so daß diese für Suchen und Verweisen genutzt werden kann.

d) Retrieval:
Neben den Browsern gibt es auch spezielle Programme, die der Suche in SGML-Dokumenten bzw. weitergefaßt der analytischen Verarbeitung dieser Dokumente unter Auswertung ihrer Struktur dienen. Vorgestellt wurden Beispiele, die mit TUSTEP realisiert waren.

e) Formatter:
Diese Programme bringen SGML-Dokumente zum Druck, ohne daß typographische Handarbeit notwendig ist. Vorgestellt wurde in dieser Funktion TUSTEP.

7 Schreibsystemvereinbarung

Von den verschiedenen Spezialthemen der TEI wurde die Writing Systems Declaration behandelt. Hier geht es darum, die Verwendung von nicht-lateinischen Schreibsystemen in leicht handhabbarer Transkription eindeutig zu dokumentieren, zu jedem Transkriptionszeichen die Bedeutung, den UNICODE-Character und das entsprechende Glyph aus dem Glyphregister zuzuordnen. Die Schreibsystemvereinbarung der TEI erfolgt in einer Form, die auch von (zukünftiger) Software ausgewertet werden kann.

8 Informationsquellen

Informationen über SGML und TEI erhält man im WWW unter folgenden URLs. Homepage der Text Encoding Initiative:
`http://www-tei.uic.edu/orgs/tei`
Die Richtlinien der TEI im Zugriff mit dem DynaWeb-Browser, der auch eine Volltextsuche erlaubt:
`http://etext.virginia.edu/TEI.html`
Webpage, in der alles Wissenswerte zu SGML gesammelt ist: `http://www.sil.org/sgml/sgml.hunl`
Steve Peppers *Whirlwind Guide to SGMLTools*, der ausführlich über SGML-Software informiert, ist via ftp erhältlich:
`ftp.falch.no/pub/SGML-Tools/sgmltool.Ut`
Eine E-mail-Diskussionsgruppe gibt es unter der Adresse:
`tei-l@uicvm.bitnet`

In einer der folgenden Hefte des LDV-Forums wird die TEI genauer vorgestellt.

Winfried Bader
Universität Tübingen Zentrum
für Datenverarbeitung
Brunnenstraße 27
D-72074 Tübingen
bader@zdv.uni-tuebingen.de

GLDV Herbstschule '96

Herausforderungen für die Computerlinguistik: Multilingualität, Semantik, Diskurs
23. - 27. September 1996
an der Otto-von-Guericke Universität Magdeburg

Kursangebot: (je Kurs 5 * 1,5 Stunden):

Elisabeth Andre (DFKI Saarbrücken):
Intelligente muttimediale Benutzerschnittstellen
John Bateman (GMD Darmstadt):
Multilinguale *Sprachgenerierung* für *Informationssysteme*
Hans Haller (IAI Saarbrücken):
Kontrollierte Sprache und *Tools*
Roland Hausser (Universität Erlangen- Nürnberg):
Semantisches Parsing
Chris Mellish (Universität Edinburgh):
Angewandte automatische Generierung von *Texten und Hypertexten* Dietmar
Rösner (Universität Magdeburg):
Wissensrepräsentation mit terminologischen Logiken

Wir versuchen, vor allem für Studenten aus Osteuropa Stipendien vergeben zu können. Bitte senden Sie Ihre Bewerbung an uns.

Frühanmeldungen bis 30. Juni 1996:

	Mitglieder der *GLDV*	*Nichtmitglieder*
Studenten:	80,- DM	100,- DM
Sonstige:	120,- DM	150,- DM

Anmeldungen nach dem 30. Juni 1996:

	Mitglieder der *GLDV*	*Nichtmitglieder*
Studenten:	120,- DM	160,- DM
Sonstige:	140,- DM	200,- DM

Weitere Informationen, mit u.a. einer Beschreibung der einzelnen Kurse, Übernachtungsmöglichkeiten, Wegbeschreibungen und einem Anmeldeformular werden auf der Homepage der Herbstschule veröffentlicht:

http://www-ai.cs.uni-magdeburg.de/herbstschule96.html

Wenn Sie ein Bett in der Jugendherberge reservieren wollen, vermerken Sie dies bitte im Anmeldeformular.

Otto-von-Guericke Universität Magdeburg
Institut für Informations- und Kommunikationssysteme
Prof. Dr. Dietmar Rösner
Universitätsplatz 2
D-39106 Magdeburg
tel: +49/391/67-1 87 18
fax: +49/391/67-1 20 18
email: herbstsehule@iik.es.uni-magdeburg.de
www: http://www-ai.es.uni-magdeburg.de/herbstsehule96.html

ECAI-96 Workshop

Corpus-oriented Semantic Analysis

Budapest, 12 or 13 August 1996

In Computational Linguistics, the semantics of natural languages is usually investigated from a rat her formal perspective. Though these studies have yielded a number of interesting results for several semantic phenomena, it is not clear whether such formal semantic theories can yet be used to automatically process real corpora. For one thing, it is often not clear whether solutions proposed for particular phenomena can be straightforwardly integrated to cope with sentences/utterances containing several phenomena. Moreover, semantic theories are often based on higher-order formal logics and do not explicitly address the problem of efficient implementations.

Existing systems for processing large corpora, on the other hand, tend to more or less neglect semantic issues. At best, such systems use a thesaurus to overcome the most drastic shortcomings in simple keyword-based approaches (e.g. for automatic text retrieval) .

The workshop aims at gathering the various strands of works which claim to obtain reasonable semantic depth while remaining, at least in theory, tractable when dealing with realistic corpora. This is only possible when the corpora concern restricted semantic domains, and when the domain knowledge has been efficiently represented. The success criterion should be the ability to entertain "intelligent dialogues" about the content of the texts. We thus use the term 'semantics' in a rather broad sense, comprising aspects which are also addressed in message or content extraction.

It is not necessary to have an implemented system, nor to be in the process of building one, to participate to the workshop. But participation will be limited to those who propose algorithmically effective solutions that can be integrated in a dialogue architecture. We are also interested in the application of standard techniques in large-corpora processing to semantic phenomena.

Important Dates:
Deadline for submission: April 1, 1996
Notification of authors: May 7, 1996 Final
versions due for: May 30, 1996 FTP versions
available: June 10, 1996 Workshop: August
12 or 13, 1996

Submissions (about 8 pages long) should be sent to Joachim Quantz. Electronic submission (preferably postscript or self-contained LATEXfiles) is strongly encouraged.

Note: Everyone attending the workshop will be required to register for the main conference. (http://wwwis.cs.utwente.nl:8080/mars/ECAI96.html)

Anybody wishing to attend the workshop without presenting a paper should contact the organizers as early as possible, since attendance will, of necessity, be limited,

Organizing Committee:

Daniel Kayser
Laboratoire d'Informatique de Paris-Nord
Universite de Paris-Nord dk@ura1507.univ-paris13.fr
Francois Levy
Laboratoire d'Informatique de Paris-Nord
Universite de Paris-Nord fl@ura1507.univ-paris13.fr
Günter Neumann
DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken
Tel. +49 681 302 5298
J. Joachim Quantz (primary contact) Technische Universität Berlin
Projekt KIT-VMll, FR 5-12
D-10587 Berlin
Tel. +49 30314 25494

Zweiter Aufruf zur Teilnahme/Second Call for Papers
KONVENS 96
3. Konferenz "Verarbeitung natürlicher Sprache"
7. - 9. Oktober 1996
Universität Bielefeld

Datum des 2. Aufrufs: 6 Februar 1996
Weitere Informationen: http://peel.lili.uni-bielefeld.de/konvens96/
Veranstalter/Organizers:

ÖGAI - Österreichische Gesellschaft für Artificial Intelligence
DGfS - Deutsche Gesellschaft f. Sprachwissenschaft/Sektion Computerlinguistik GI - Gesellschaft f. Informatik, FA 1.3 "Natürliche Sprache"
GLDV - Gesellschaft f. linguistische Datenverarbeitung
ITG/DEGA - Informationstechnische Gesellschaft/Deutsche Ges. f. Akustik

Tagungsleiter / chairmen:

Dafydd Gibbon & Henning Lobin, Universität Bielefeld

Programmkomitee/program committee:

Ernst Buchberger, ÖGAI (U Wien)
Stefan Busemann, GI (DFKI Saarbrücken) Dafydd Gibbon, lokal
Wolfgang Hoeppner, GI (U Duisburg) Rüdiger Hoffmann, ITG/DEGA (TU Dresden) Tibor Kiss, DGfS (IBM Deutschland, WHZ) Winfried Lenders, GLDV (IKP Bonn) Henning Lobin, lokal
Ute Seewald, GLDV (U Hannover)
Harald Trost, ÖGAI (U Wien)
Hans Uszkoreit, DGfS (U Saarbrücken)

Die Gesellschaften DGfS, GI, GLDV, ITG/DEGA und ÖGAI veranstalten alle 2 Jahre gemeinsam eine Tagung zum Thema Verarbeitung natürlicher Sprache unter der Bezeichnung 'KONVENS'. Die KONVENS 96 wird vom 7.-9. Oktober 1996 in Bielefeld stattfinden.

Die KONVENS soll einen Querschnitt über die aktuelle Forschung auf allen Gebieten der Verarbeitung natürlicher Sprache am Computer zu bieten. Dazu ist Mitarbeit aus allen beteiligten Disziplinen, wie Informatik, Künstliche Intelligenz, Linguistik, Philosophie, Psychologie und Nachrichtentechnik erwünscht. Es sollen sowohl grundlagenorientierte Forschungsaspekte und -resultate vertreten sein als auch innovative Anwendungen. Bevorzugt werden Papiere über erfolgreich durchgeführte und implementierte Vorhaben.
Als Schwerpunktthema für die KONVENS 96 wurde

Multimediale und multimodale Sprachverarbeitung

gewählt. Zum Schwerpunktthema werden eingeladene Vorträge und Einführungskurse, die der Tagung vorangehen, angeboten. Hauptvorträge werden im Kontext des Tagungsthemas von

Prof. Dr. Wolfgang Wahlster (DFKI, Saarbrücken), über die wissenschaftlichen Ergebnisse der ersten Phase von VERBMOBIL

Dr. Julia Hirschberg (AT&T Media Laboratory), über Aspekte der Diskursverarbeitung in gesprochener Sprache gehalten. Die Berücksichtigung des Rahmenthemas in den Beiträgen ist erwünscht aber nicht verbindlich.

Erbeten werden Beiträge, Posters und Systemvorführungen zu allen Gebieten der Verarbeitung natürlicher Sprache (nicht nur zum Schwerpunktthema), insbesondere aber zu folgenden Themen:

> Psycholinguistische Aspekte multimodaler Sprachverarbeitung t> Integration visueller und akustischer Sprachverarbeitung
> Multimodale Sprache in Hypermedien und Hypertext
> Repräsentationsformalismen für Language und Speech t> Integrative und hybride Ansätze

> Ressourcen und Standardisierung in der Sprach- und Textverarbeitung r>

Werkzeuge für Sprach- und Texttechnologien

> Anwendungen multimedialer und multimodaler Sprachverarbeitung

Auf der KONVENS 96 werden erstmalig W 0 R K S H 0 P S vorgesehen.

Einreichen von *Beiträgen/Submission* of *papers*

> Beiträge werden anonym begutachtet

> Jedem Beitrag muß ein gesondertes Deckblatt beigelegt werden, auf dem Name und Affiliation der Verfasser sowie Titel und Art des Beitrags aufgeführt werden, um die. anonyme Begutachtung zu ermöglichen

> Beiträge sollen in fünf Papier-Exemplaren, DIN A4, Times 12pt, sowie in elektronischer Form per email (vorzugsweise 17\TEXoder PS) eingereicht werden

> Formate:

VORTRAGSBEITRÄGE sollen 10 Seiten nicht überschreiten

POSTERBEITRÄGE sollen max. 4 Seiten umfassen

WORKSHOP-ANMELDUNGEN sollen max. 6 Seiten umfassen, die Bedeutung des Themas begründen und einen Überblick über die voraussichtlichen Teilnehmer und deren Beiträge angeben

> Ankündigungen von Systemvorführungen müssen Titel und Kurzbeschreibung sowie die benötigte technische Infrastruktur spezifizieren.

> Alle Beiträge sollen eine Zusammenfassung auf Deutsch und Englisch von jeweils max. 12 Zeilen enthalten.

> Tagungssprachen sind D e u t s c h und Eng I i s c h.

> Alle Einreichungen werden von mindestens 2 unabhängigen Gutachtern beurteilt, die vom Programmkomitee bestimmt werden.

> Der Tagungsband, in dem alle angenommenen Beiträge und Posterbeiträge abgedruckt werden, wird rechtzeitig zur Tagung erscheinen.

> Beiträge und Workshop-Vorschläge sollen an Dafydd Gibbon (Adresse unten) geschickt werden.

Termine/Deadlines

01.3.96	Einreichung von Workshopthemen
15.3.96	Benachrichtigung über Workshop-Auswahl
15.4.96 -	Einreichung von Beiträgen und Posterbeiträgen
15.5.96	Benachrichtigung über Annahme oder Ablehnungen
	Einsendeschluß für Workshop-Unterlagen durch Veranstalter Abgabe
15.6.96	der druckfertigen Vorlagen für den Tagungsband Anmeldung von
15.7.96	Systemvorführungen

	Organisation:	Tagungsprogramm und Einreichung der Beiträge:
	Dr. Henning Lobin	Prof. Dr. Dafydd Gibbon
e-mail:	lobin@lili.uni-bielefeld.de	gibbon@spectrum.uni-bielefeld.de
Tel.:	+49.521.106.3679	+49.521.106.3509
Fax:	+49.521.106.2996	+49.521.106.6008

Universität Bielefeld Fakultät
für Linguistik und
Literaturwissenschaft
Postfach 100131
D-33501 Bielefeld

WWW: http://peel.lili.uni-bielefeld.de/konvens96/

Anläßlich der KONVENS '96
findet ein Workshop zu folgendem Thema statt:
Präpositionalsemantik und PP-Anbindung

Zur KONVENS und zum Workshop gibt's bereits folgende WWW-Seiten:
http://coral.lili.uni-bielefeld.de/konvens96/info.html bzw.
http://voss.fernuni-hagen.de/gebiete/pi7/workshop.html

Workshop- Organisatoren:

Stephan Mehl
Uni Duisburg, Computerlinguistik, 47048 Duisburg
Tel.: 0203/3792876, Email: he234me@unidui.uni-duisburg.de

Andreas Mertens
FernUni Hagen, Praktische Informatik VII/KI, 58084 Hagen Tel.:
02331/9874525, Email: andreas.mertens@fernuni-hagen.de

Marion Schulz
Fern Uni Hagen, Praktische Informatik VII/KI, 58084 Hagen Tel.:
02331/9874524, Email: marion.schulz@fernuni-hagen.de

Workshop- Beschreibung:
Präpositionen sind hochgradig mehrdeutig. So kann z.B. "über" lokal (über der Brücke kreisen), direktional (über die Brücke fahren), thematisch (über die Brücke reden, über Ostern reden) oder temporal (über Ostern wegfahren) interpretiert werden. Dieselbe Präpositionalphrase (PP) kann daher in verschiedenen Kontexten unterschiedliche semantische Rollen einnehmen. (Hinzu kommen semantische Unterschiede, die nicht zu verschiedenen Rollenzuweisungen führen; z.B. kann "vor" ("vor dem Auto") in Relation zum Sprecher oder in Relation zum Objekt interpretiert werden. Diese Problematik soll im folgenden ausgeklammert werden.).

Wie die Beispiele zeigen, tragen zur semantischen Disambiguierung von Präpositionen bzw. zur funktionalen Disambiguierung von PPs verschiedene Faktoren bei:

> morphologische (Rektion von Dativ vs. Akkusativ);

> syntaktische (Festlegung durch den Valenzrahmen wie bei "reden");
> semantische bzw. enzyklopädische (Kompatibilität der Präposition mit dem regierten Nomen, Kompatibilität der gesamten PP mit dem Matrixwort);

> pragmatische (gesamter Äußerungszusammenhang, z.B. bei "Ich rede mit ihr über Ostern").

Die Disambiguierung kann also unterschiedlich schwierig sein.

Neben ihrer semantisch-funktionalen Ambiguität ergeben sich bei der syntaktischen Analyse von Präpositionalphrasen häufig strukturelle Mehrdeutigkeiten. Zwar ist dieses Phänomen von dem der lexikalischen Mehrdeutigkeit der Präpositionen unabhängig: Es gibt zahlreiche Sätze mit eindeutiger Präposition, aber unklarer Anbindung (z.B. "Er holt. seine Tante aus Berlin ab." [d.h. vom Berliner/Düsseldorfer Flughafen]); der umgekehrte Fall ist wesentlich seltener, kommt aber auch vor ("Er steht seit Jahren auf Korkparkett." [befindet sich darauf/schwärmt dafür]). Häufig treten beide Phänomene aber auch gemeinsam auf (z.B. in "Sie hat den Landstreicher mit dem Schäferhund vertrieben." [mit Hilfe/in Begleitung]).

Für die strukturelle Disambiguierung von PPs gilt dasselbe wie für die lexikalische Disambiguierung von Präpositionen: es können Regeln aus den verschiedensten Beschreibungsebenen eine Rolle spielen. Man beachte, daß die o. g. Beispiele global mehrdeutig sind, d.h. ohne Kenntnis des Äußerungszusammenhangs nicht disambiguiert werden können. Dies ist der schwierigste, aber auch seltenste Fall. Die meisten Präpositionen sind aber zum Zeitpunkt ihrer Verarbeitung zunächst einmal lokal mehrdeutig (syntaktisch wie semantisch), d.h. die zu ihrer Disambiguierung benötigte Information folgt erst später.

Die Architektur sprachverarbeitender Systeme wird damit vor ein dreifaches Problem gestellt:

> praktisch jede Präposition ist beim Einlesen semantisch und/oder von ihrer syntaktischen Zuordnung her mehrdeutig;

> zur Disambiguierung ist syntaktisches und/oder semantisch/enzyklopädisches Wissen und/oder pragmatisch/referentielles Wissen erforderlich;

> dieses Wissen wird evtl. erst später bekannt.

Thema des Workshops sollen mögliche Lösungen dieses Architekturproblems sein:

> Werden mehrere Möglichkeiten gleichzeitig verfolgt, oder wird eine ausgewählt? Wenn letzteres, nach welchen Kriterien?

> Werden syntaktische und semantische Analyse seriell oder parallel angeordnet?

Lösungsvorschläge können einerseits auf kognitionswissenschaftlichen Untersuchungen beruhen. Andererseits sind auch Erfahrungsberichte aus dem praktischen Einsatz implementierter Systeme erwünscht. Neben den o. g. Fragen kommt dann hinzu:

> Welche Wissensbasen wurden zur Disambiguierung eingesetzt, und mit welchem Er folg?

Die Organisatoren werden den Workshop durch einen Abriß des Problembereichs und eine Synopse einiger möglicher Lösungsansätze einleiten. Die Teilnehmer erhalten vorab Positionspapiere der Referenten, so daß einerseits die Beiträge aufeinander bezogen werden können und andererseits eine fruchtbare Diskussion möglich wird.
Alle KONVENS- Teilnehmer sind herzlich willkommen!

SIGIR '96

Call for Papers

19th International Conference on Research and Development in Information Retrieval

Held at ETH, Zurich, Switzerland

August 18 - August 22, 1996

Organized by:

UBILAB, Union Bank of Switzerland, Zurich, Switzerland ETH, Swiss Federal Institute of Technology, Zurich, Switzerland

In co-operation with:
ACM
AICA-GLIR (Italy)
BCS-IRSG (UK)
CEPIS-EIRSG (Europe)
GI (Germany)
IPSJ (Japan)
OCG (Austria)
SI (Switzerland)

About the conference
SIGIR '96 is an international research conference on information retrieval (IR) theory, systems, and applications. The ACM SIGIR conference occurs annually, alternating between locations in N orth America and elsewhere (e.g. Europe). It attracts a broad range of professionals including theoreticians, developers, publishers, researchers, educators, and designers of systems, interfaces, information bases, and related applications.

Call for papers
SIGIR '96 seeks original papers (i.e. never before submitted or published) on significant contributions to the broad field of information storage and retrieval covering the handling of all types of information as well as the underlying theories, models, and implementations.

We encourage discussions of experimental studies, tests of usability, explorations of information retrieval behavior, reports on largescale system performance, and demonstrations of advanced approaches. We prefer theoretical contributions to have sufficient proof of utility, and designs that have been proven by a prototype. Similarly, reports on smallscale experiments should include convincing arguments or simulations to show their likelihood of generalization.

Topics

Topics are addressed at the theoretical, experimental and applications level and include, but are not limited to:

Information Retrieval Theory: Probabilistic and Logic Retrieval Models, Data Fusion.

User Interaction: Models, Interfaces, Query (Re-) Formulation, Visualization.

Hypermedia: Audio, Video, and Image Retrieval, Links, Composite Documents.

Experimentation: Test Collections, Evaluation Measures.

Natural Language Processing: Indexing and Retrieval using Linguistic Resources, Multilingual Retrieval.

Systems and Implementation Issues: Interaction with Database Systems, Networked Systems, Compression, Emcient Query Evaluation.

Applications: Workflow Management, Electronic Publishing, Digital Libraries.

Instructions for contributors

Submissions to SIGIR '96 can be made in the form of papers, tutorials, panels, posters, or demonstrations, and should be addressed to the relevant Chairs. With the exception of papers and posters, submissions may be made via e-mail (plain ASCII text). All submissions should include complete contact information including mail address, telephone, fax, and e-mail.

Papers

Papers (4 copies) should be submitted in English to the relevant Program Co-chair. Papers should contain at most 5000 words and should be double-spaced. The first page must contain the title of the paper and an abstract of not more than 100 words, but
no indication ab out the author(s) and affiliation(s). In addition, authors must provide a
separate page with the title, the author name(s), and the author affiliation(s), plus complete contact information (mailing address, telephone, fax, and e-mail) for the author to whom correspondence should be sent. Please indicate if the paper is to be considered for the Best Student Paper A ward. This A ward requires that the first and primary author be a fulltime student at time of submission.

Tutorials

SIGIR '96 will begin with a full day of tutorials, each of which is intended to cover a single topic in detail. Proposals are solicited for tutorials of either a half day (4 hours) or full day in length. Submissions should be made to the Tutorials and Panels Chair and should include an extended abstract (3 - 5 pages) and an indication of the tutorial length, objectives, and intended audience.

Panels

SIGIR '96 will include a small number of panel sessions. These are intended to ex amine issues of interest to the research and development community and to stimulate lively debate between panelists and audience. Proposals are solicited from prospective moderators. Submissions to the Tutorials and Panels Chair should include an extended abstract outlining the proposed topic and a list of panel members.

Demonstrations

Demonstrations provide an opportunity far first-hand, interactive experience with information retrieval systems. Proposals (up to 3 pages) should be submitted to the

Demonstrations Chair. The proposal should indicate how the demonstration will illustrate new ideas and describe the technical specifications of the system. The hardware, software, and network requirements for the demonstration, including the electrical requirements of the equipment, should be indicated.

Posters

SIGIR '96 will include poster presentations to offer researchers an opportunity to present late-breaking results, significant work in progress, or research that is best communicated in presentational format. Abstracts of posters will appear in the conference proceedings.

Submissions to the Posters Chair should consist of an extended abstract (3 copies) of approximately 3 - 4 pages emphasizing the research problem and the methods being used. In addition, a separate cover page is required containing the title of the poster, the name and affiliation of the author(s) as well as complete contact information.

Workshops

Proposals are being solicited from both individuals and groups for one-day workshops to be held on August 22, 1996. Workshops bring together researchers to share information and discuss a topic that relates to their particular field of expertise. Submissions of up to 3 pages should be made to the Workshops Chair. They should include the theme and goal of the workshop, the planned activities, the maximum number of participants and the selection process, plus a list of potential participants. Please include a CV for each organizer detailing relevant qualifications and experience. After the workshop, organizers are to provide an article summarizing the workshop for SIGIR Forum.

Important dates

Presently: E-mail sigir96@ubilab.ubs.ch with subject 'subscribe' and an empty message body to be added to the mailing list.

January 26, 1996: Submission of papers to the relevant Program Co-chair.

March 1 1996: Submission of proposals for tutorials, panels, demonstrations, posters, and workshops to the relevant Chair.

April 3, 1996: Author notification.

May 10, 1996: Final manuscript due in camera ready and electronic forms.

Further information

Detailed information on the SIGIR '96 conference is available on the World Wide Web. URL: http://www.ubilab.ubs.ch/sigir96/welcome . html. Questions should be addressed to sigir96@ubilab.ubs.ch.

General Conference *Chair:*
Hans- Peter Frei
UBILAB, Union Bank of Switzerland Bahnhofstr. 45,
8021 Zurich, Switzerland sigir96@ubilab.ubs.ch
Phone: +41/1-2365714; Fax: +41/1-236 4671
Local Arrangements Chair:
Gabriele Sonnenberger
UBILAB, Union Bank of Switzerland
P.O. Box 2336, 8033 Zurich, Switzerland
sigir96@ubilab.ubs.ch
Phone: +41/1-2344457; Fax: +41/1-2346518 *Local Organization:*
Sabina Braeuer

UBILAB, Union Bank of Switzerland
P.O. Box 2336, 8033 Zurich, Switzerland
sigir96@ubilab.ubs.ch
Phone: +41/1-234 2147; Fax: +41/1-234 6518

Treasurer:

Peter Schaeuble
Informationssysteme, ETH Zentrum
8092 Zurich, Switzerland
sigir96@ubilab.ubs.ch
Phone: +41/1-632 7222; Fax: +41/1-632 1172

Program Co-chairs:

Europe, Africa:
Peter Schaeuble
Informationssysteme, ETH Zentrum
8092 Zurich, Switzerland
schauble@inf.ethz.ch
Phone: +41/1-632 7222; Fax: +41/1-632 1172

North and South America:
Donna Harman
National Institute of Standards and Technology
Building 225, Room A216, Gaithersburg, MD 20899, USA
harman@potomac.ncsl.nist.gov
Phone: +1/301-975 3569; Fax: +1/301-840 1357

Asia, Pacific:
Ross Wilkinson
RMIT
723 Swanston Street, Carlton, 3053, Australia
ross@kbs.citri.edu.au
Phone: +61/3-9282 2485; Fax: +61/3-9282 2490

Tutorials and Panels Chair:

Norbert Fuhr
Lehrstuhl Informatik VI, Universitaet Dortmund
44221 Dortmund, Germany
fuhr@ls6.informatik.uni-dortmund.de
Phone: +49/231-755 2045, or 2779; Fax: +49/231-755 2405

Posters Chair:

Elizabeth D. Liddy
4-206 Center for Science and Technology, Syracuse University,
Syracuse, NY 13210, USA
liddy@mailbox.syr.edu
Phone: +1/315-443 4456; Fax: +1/315-443 5806

Demonstrations Chair:

Marc Rittberger
Informationswissenschaft, Universitaet Konstanz
P.O. Box 5560, 78434 Constance, Germany
Marc.Rittberger@uni-konstanz.de
Phone: +49/7531-88 3595; Fax: +49/7531-88 26 01

Workshops Chair:

Robert R. Korfhage
Dept. of Information Science, Univ. of Pittsburgh
Pittsburgh, PA 15260, USA
korfhage@lis.pitt.edu
Phone: +1/412-624 9420 or: 421 1428; Fax: +1/412-624 2788

Program Committee:

IJsbrand Jan Aalbersberg, Philips, The Netherlands
Maristella Agosti, U. Padua, Italy
Peter Anick, Digital Equipment Corp.,USA
Richard K. Belew, UC San Diego, USA
Nicholas Belkin, Rutgers U., USA
Abraham Bookstein, U. Chicago, USA
Giorgio Brajnik, U. Udine, Italy
Peter D. Bruza, QUT, Australia
Chris Buckley, Cornell U., USA

Yves Chiaramella, LGI-IMAG, France
Stavros Christodoulakis, Technical U. of Crete, Greece
W. Bruce Croft, U. Massachusetts,USA
Raya Fidel, U. of Washington, USA
Edward Fox, Virginia Tech, USA
Norbert Fuhr, U. Dortmund, Germany
Richard Furuta, Texas A&M U., USA
Frederic Gey, UC Berkeley, USA
Gregory Grefenstette, Rank Xerox Research Centre, France
Micheline Hancock-Beaulieu, City University, UK
David Harper, Robert Gordon U., UK
William Hersh, Oregon Health Sciences U. USA
David Hull, Rank Xerox Research Centre, France
Nancy Ide, Vassar College, USA
Peter Ingwersen, School of Librarianship, Denmark
Tetsuya Ishikawa, ULIS, Japan
Kalervo Jarvelin, U. Tampere, Finland
Paul Kantor, Rutgers U., USA
Haruo Kimoto, NTT, Japan
Shmuel T. Klein, Bar-Ilan U., Israel
Robert Korfhage, U. Pittsburgh, USA
Donald Kraft, Louisiana State U., USA
Dik L. Lee, HongKong U., HongKong
Joon Ho Lee, KIST, Korea
David Lewis, AT&T, USA
Elizabeth D. Liddy, Syracuse U., USA
Elke Mittendorf, ETH, Switzerland
Alister Moffat, U. of Melbourne, Australia
Sung Hyun Myaeng, Chungnam National U., Korea
Desai Narasimhalu, National U. of Singapore, Signapore
Yasushi Ogawa, RICOH, Japan
Jan O. Pedersen, Xerox PARC, USA
Annelise Mark Pejtersen, Risoe, Denmark
Keith van Rijsbergen, U. Glasgow, UK
Stephen Robertson, City U., UK
Tefko Saracevic, Rutgers U., USA
Fabrizio Sebastiani, IEI-CHR, Italy
Alan Smeaton, Dublin City U., Ireland
Phil Smith, Ohio State U., USA
Craig Stanfill, Ab Initio Software, USA
Jean Tague-Sutcliffe, U. of Western Ontario, Canada
Ulrich Thiel, GMD IPSI, Germany
Richard Tong, Verity, USA
Howard Turtle, West Publishing, USA
Ellen Voorhees, Siemens, USA
Peter Willett, U. Sheffield, UK

Aus der Arbeit von Vorstand und Beirat

Tagungsband
Der Tagungsband der Jahrestagung der GLDV 1995 in Regensburg ist inzwischen unter dem Titel "Angewandte Computerlinguistik" im Olms-Verlag, Hildesheim erscheinen.

Neues aus den Arbeitskreisen
AK-Ausbildung und Berufsperspektiven
Leitung: Christian Wolff
Der Arbeitskreis "Ausbildung und Berufsperspektiven", dessen Leitung Herr Christian Wolff aus Leipzig übernommen hat, hat sich nach der Jahrestagung in Regensburg neu konstituiert und ist inzwischen zu einem Treffen zusammengekommen. Herr Wolff bittet um Mitteilung der Änderungen, die in der Zwischenzeit in die Studienordnungen der Studiengänge aufgenommen wurden, die in der GLDV-Publikation "Studienführer Linguistische Datenverarbeitung/Computerlinguistik für die deutschen Universitäten" von 1991 verzeichnet sind. Herr Wolff beabsichtigt, eine WWW-Seite einzurichten, auf der der Studienführer auch elektronisch verfügbar gemacht werden soll Möglicherweise kann diese WWW-Seite außerdem dazu genutzt werden, Angaben über Trends in der Computerlinguistik zu integrieren, wo über Einschreibungszahlen, Studienzahlen und eventuell auch Stellenausschreibungen Auskunft gegeben wird.

AK-Parsing in Morphologie und Syntax
Leitung: Roland Hausser
Die Dokumentation der 1. MORPHOL YMPICS, die im März 1994 in Erlangen vom Arbeitskreis "Parsing in Morphologie und Syntax" unter der Leitung von Roland Hausser durchgeführt wurde, erscheint im Frühjahr 1996 im Niemeyer-Verlag in der Reihe Sprache und Information unter dem Titel "Linguistische Verifikation".

AK-Hypermedia
Leitung: Angelika Storrer
Der neu gegründete Arbeitskreis "Hypermedia" unter der Leitung von Angelika Storrer veranstaltete am 21./22. März 1996 am IDS in Mannheim einen Workshop zum Thema "Hypermedia für Lexikon und Grammatik".

KONVENS '96
Die GLDV hat als Mitglieder des Programmkomittees der KONVENS '96 Winfried Lenders und Uta Seewald benannt. Auf der diesjährigen KONVENS werden erstmals Workshops angeboten. Themenvorschläge hierzu werden auch aus den Reihen der GLDV unterbreitet.

Einladung zur GLDV-Mitgliederversammlung 1996

Die Mitgliederversammlung der GLDV für 1996 findet in diesem Jahr während der KONYENS-Tagung in Bielefeld statt, und zwar am Dienstag, dem 8. Oktober 1996, 18 Uhr. Der Sitzungsraum wird in der endgültigen Einladung noch mitgeteilt.

Vorläufige Tagesordnung:

1. Eröffnung der Sitzung, Begrüßung der Teilnehmer und Regularien
2. Endgültige Festlegung der Tagesordnung
3. Berichte des Vorstands mit Kassenbericht und Bericht der Kassenprüfer
4. Haushaltsplan 96/97
5. Entlastung des Vorstands
6. Wahl von Kassenprüfern
7. Bericht des Beirats
8. Bericht aus den Arbeitskreisen
9. Jahrestagung 1997 und 1998
10. Arbeitsprogramm 1996/97
11. Verschiedenes

Weitere Tagesordnungspunkte können seitens der Mitglieder gern. § 17 der Satzung bis zum 30. September 1996 beim Vorstand beantragt werden.
Die endgültige Einladung ergeht rechtzeitig an alle Mitglieder.
Der Vorstand ruft dingend zur Teilnahme der Mitglieder nicht nur an der Mitgliederversammlung, sondern an der gesamten Konvens auf.

Protokoll

der Mitgliederversammlung der GLDV vom 30.03.1995 in Regensburg

Beginn: 16.30 Uhr Ende: 18.30 Uhr Sitzungsleitung: Winfried Lenders

Tagesordnung

TOP 1: Regularien
TOP 2: Bericht des Vorstands und Kassenbericht 1994
TOP 3: Entlastung des Vorstands
TOP 4: Wahl von Kassenprüfern
TOP 5: Bericht des Beirats
TOP 6: Bericht der Arbeitsgruppen/-kreise
TOP 7: Zusammensetzung und Wahl des Beirats:
Diskussion und Beschlußfassung zu einer vorgeschlagenen Änderung von § 13, Abs. (2) der Satzung
TOP 8: Neuwahlen
8.1: Wahl eines Wahlvorstandes
8.2: Kandidatenliste: Vorstand
8.3: Kandidatenliste: Beirat
TOP 9: Arbeitsprogramm 1995/96
TOP 10: Nächste Jahrestagungen
TOP 11 : Verschiedenes

TOP 1: Regularien
Der Vorstandsvorsitzende, W. Lenders, stellt fest, daß die Tagesordnung mit der Einladung an die Mitglieder versandt und im LDV-Forum veröffentlicht worden ist. Es sind 30 Mitglieder anwesend. Die Öffentlichkeit wird von der Mitgliederversammlung zugelassen, zwei Gäste sind anwesend. Anträge auf Stimmübertragungen liegen nicht vor. Das Protokoll der letzten Mitgliederversammlung in Kiel, das den Mitgliedern durch das LDV-Forum zugegangen ist, wird genehmigt. Anträge zur Tagesordnung liegen nicht vor. Die Tagesordnung wird in der vorgelegten Form angenommen.

TOP 2: Bericht des Vorstands und Kassenbericht 1994
W. Lenders berichtet, daß der Vorstand dank des Emailsystems in regelmäßigem Gedankenaustausch steht. Darüber hinaus haben in den vergangenen zwei Jahren jeweils drei Vorstandssitzungen stattgefunden. Der Vorstand hat sich um eine Aktivierung und Motivierung der Arbeit der Arbeitskreise bemüht. Von den Aktivitäten der Arbeitskreise sind besonders hervorzuheben die Organisation und Durchführung der MORPHOL YMPICS durch den Arbeitskreis "Parsing in Morphologie und Syntax" unter der Leitung von Roland Hausser, die Veranstaltungen des Arbeitskreises "Korpora" unter der Leitung von Robert Neumann sowie das Kolloquium des Arbeitskreises "Lexikographie" unter der Leitung von Nico Weber.

W. Lenders berichtet, daß im vergangenen Jahr ein Arbeitskreis "Historisch-Vergleichende Sprachwissenschaft" unter der Leitung von Prof. Gippert von der Universität Frankfurt eingerichtet wurde. Der Vorstandsvorsitzende beklagt, daß sich trotz des Werbens für die Neuauflage der Broschüren "Studienführer Linguistische Datenverarbeitung/Computerlinguistik für die deutschen Universitäten" und "Ausbildungsprofile von CL-Lehrangeboten" bislang kein Leiter für den Arbeitskreis "CL-Studiengänge/Berufsperspektiven" gefunden hat.

Die Aufgaben des Vorstands umfaßten zuletzt die Vorbereitung der Jahrestagung in Regensburg sowie die Planung der GLDV-Herbstschule, die abweichend von der ursprünglichen Planung nun in Bonn unter dem Thema "Modeme Methoden der Corpusanalyse" stattfinden wird. Der Vorstandsvorsitzende ruft die Mitglieder auf, für die Herbstschule zu werben und macht auf die in Regensburg ausgelegten Faltblätter aufmerksam. Der Vorsitzende berichtet, daß eine Bereinigung der Mitgliederliste stattgefunden hat, aus der solche als Mitglieder geführten Personen gestrichen wurden, die seit mehreren Jahren in ihren Beitragszahlungen säumig sind. Danach beläuft sich die Zahl der GLDV-Mitglieder derzeit auf 315. Die Mitgliederzahl ist etwa konstant geblieben. 24 Beitritten stehen seit August 1993 26 Austritte gegenüber. Der Schatzmeister R. Hausser legt den Kassenbericht des Haushaltsjahres 1994 und den Haushaltsplan 1995 vor. Die Summe der Einnahmen belief sich 1994 (1.1.94 - 8.12.94) auf 26.127,63 DM. Am 8.12.94 betrug der Kontostand 9.847,41 DM; aktuell beträgt er 7.431,16 DM. R. Hausser berichtet, daß die Beiträge für das laufende Haushaltsjahr im August eingezogen werden. Die zu erwartenden Einnahmen belaufen sich auf 23.900 DM und decken die für 1995 angesetzten Ausgaben.

Die Kasse wurde von Herrn Lutz und Herrn Schröder am 3.3.1995 geprüft. Herr Lutz trägt vor, daß die Prüfung den Zeitraum vom 1.1.94 bis 8.12.94 umfaßte. Es lag eine Ein- und Ausgabenrechnung, das Journal '94, die Kontoauszüge sowie Belege für Ein- und Ausgaben zur Prüfung vor. Herr Lutz beantragt die Entlastung des Vorstands (siehe TOP 3). Die Kassenprüfer sprechen zwei Empfehlungen an den Vorstand aus: zum einen wird empfohlen, die Ausgaben und Einnahmen für das LDV-Forum künftig detaillierter aufzuschlüsseln und die einzelnen Positionen aussagekräftig zu spezifizieren. Zum anderen sollen Reisekosten im Zusammenhang mit Vorstands- und Beiratssitzungen, die eine Höhe von 500,- DM übersteigen, dem Vorstand zur Genehmigung vorgelegt werden. Es wird der Antrag gestellt, die Mitgliederversammlung möge beschließen, diese Empfehlungen zur Abstimmung zu bringen. Dem Antrag wird mit einem Stimmenverhältnis von 14:5:4 Jastimmen:Neinstimmen:Enthaltungen zugestimmt. Für die Einhaltung der von den Kassenprüfern an den Vorstand ausgesprochenen Empfehlungen stimmen 20 Mitglieder bei 6 Enthaltungen und 1 Gegenstimme.

TOP 3: Entlastung des Vorstands

Nach erfolgtem Bericht der Kassenprüfer (vgl. TOP 2) stimmt die Mitgliederversammlung einstimmig für die Entlastung des Vorstands. Nach erfolgtem Bericht der Kassenprüfer (vgl. TOP 2) stimmt die Mitgliederversammlung einstimmig für die Entlastung des Vorstands.

TOP 4: Wahl von Kassenprüfern

Herr Lutz und Herr Schröder werden zur Fortsetzung ihres Amtes als Kassenprüfer vorgeschlagen. Die Mitgliederversarnmlung stimmt der Nominierung von Herrn Lutz und Herrn Schröder einstimmig zu. Beide nehmen die Wahl an (eine entsprechende schriftliche Erklärung von Herrn Schröder liegt vor).

TOP 5: Bericht des Beirats

Herr Knorz berichtet, daß bei den drei 1994 stattgefundenen gemeinsamen Sitzungen von Vorstand und Beirat nur ein bzw. maximal drei Beiratsmitglieder anwesend waren. Auch ein in Aussicht gestellter Reisekostenzuschuß blieb hier ohne Wirkung. G. Knorz stellt fest, daß sich die Sitzungen von Vorstand und Beirat fruchtbar auf das Zustandekommen des LDV-Forums ausgewirkt haben. Der Vorschlag, studentische Arbeiten bei der diesjährigen Jahrestagung in

das Programm aufzunehmen, geht unter anderem auf die Initiative des Beirats zurück. Darüber hinausgehende Aktivitäten des Beirats hat es nicht gegeben.

TOP 6: Berichte der Arbeitsgruppen/-kreise

Frau Angelika Storrer vom IDS in Mannheim regt die Einrichtung eines Arbeitskreises "Hypertext/Hypermedia" an und erklärt sich zur Übernahme der Leitung dieses Arbeitskreises bereit. Ihr schwebt vor, eine erste Veranstaltung gegebenenfalls gemeinsam mit dem Arbeitskreis "Lexikographie" zu gestalten. Herr Burghard Rieger ruft zur Mitarbeit in einem Arbeitskreis "Fuzzy Linguistik" auf. Frau Lutz-Hensel macht darauf aufmerksam, daß sie als Mitglied des Arbeitskreises "CL-Studiengänge/Berufsperspektiven" ein Plakat mit zur Diskussion gestellten Thesen in Regensburg ausgehängt hat. AK "CL-Studiengänge/Berufsperspektiven": W. Lenders berichtet, daß die Bemühungen um die Suche nach einer Nachfolge der Leitung des Arbeitskreises "CL-Studiengänge/Berufsperspektiven" bisher nicht erfolgreich gewesen sind. Da der Vorstand eine Neuauflage der Broschüren für dringend erforderlich hält, wurde bereits die Frage erörtert, diese Unternehmung in Saarbrücken anzusiedeln und dort von Hilfskräften im Lohnauftrag zu erstellen. Auch von Seiten der Mitgliederversammlung wird die Überarbeitung der Studienführer als eine der zentralen Aufgaben der Gesellschaft betrachtet. Die Mitgliederversammlung beauftragt den Vorstand, sich der Umsetzung der Überarbeitung der Broschüren anzunehmen. Frau Lutz-Hensel signalisiert ihre Bereitschaft, die Überarbeitung mit Material und Informationen zu unterstützen. 1. Krause macht darauf aufmerksam, daß in das Projekt der Überarbeitung der Broschüren künftig auch die "Blätter zur Berufskunde" einbezogen werden müssen, für deren organisatorische Betreuung er aufgrund seines fachlichen Wechsels in die Informatik fortan nicht mehr zur Verfügung stehen wird. Die Mitgliederversammlung schlägt vor, daß alle Arbeitskreise der GLDV aufgefordert werden sollen, auf der Mitgliederversammlung über ihre Aktivitäten kurz zu berichten.

TOP 7: Zusammensetzung und Wahl des Beirats

Diskussion und Beschlußfassung zu einer vorgeschlagenen Änderung von § 13, Abs. (2) der Satzung W. Lenders stellt fest, daß bezüglich der vom Vorstand vorgeschlagenen Satzungsänderung auf der Mitgliederversammlung in Wien bereits ein erstes Meinungsbild erstellt wurde. Der Entwurf der Satzungsänderung ist den Mitgliedern mit der Einladung zur Mitgliederversammlung zugegangen. Einzelne Reaktionen auf die vorgeschlagene Formulierung sind per Email eingegangen. Nach der Beseitigung der in der vorgeschlagenen Formulierung enthaltenen Inkonsistenz lautet der Text, über den die Mitgliederversammlung abzustimmen hat, wie folgt: "Der Beirat besteht aus maximal 7 Mitgliedern. Vier seiner Mitglieder werden auf die Dauer von zwei Jahren, vom Tag der Wahl an gerechnet, von den Mitgliedern gewählt. Bis zu drei weitere Mitglieder können durch den Vorstand kooptiert werden. Der Beirat bleibt bis zur Neuwahl im Amt. Wählbar sind auch natürliche Personen, die der Vereinigung nicht angehören. Vorstandsmitglieder können nicht zugleich Mitglieder des Beirats sein. Der Beirat wählt einen Beiratssprecher. Die Mitgliederversammlung stimmt mit 28 Ja-Stimmen und 2 Enthaltungen für die Änderung von § 13, Abs. (2) der Satzung.

TOP 8: Neuwahlen

W. Lenders unterrichtet die Mitgliederversammlung über das Procedere der Aufstellung von Kandidatenlisten für die Wahl von Vorstand und Beirat.

TOP 8.1: Wahl des Wahlvorstandes

Als Kandidaten für den Wahlvorstand werden Prof. Dr. Istvan Baton, Universität Koblenz-Landau, Prof. Dr. Burkhard Schaeder, Universität GHS Siegen, und Bernhard Schroeder, M.A., Universität Bonn, vorgeschlagen. Die Mitgliederversammlung nimmt die Kandidatenliste bei I Enthaltung einstimmig an.

TOP 8.2: Kandidatenliste: Vorstand

Alle Mitglieder des Vorstandes stellen sich der Wiederwahl. Die Mitgliederversammlung nimmt die Kandidatenliste bei 10 Enthaltungen einstimmig an.

TOP 8.3: Kandidatenliste: Beirat

Die den Mitgliedern mit der Einladung zugegangene Kandidatenliste für den Beirat wird um einen Kandidaten erweitert und enthält folgende Personen: Prof. Dr. Gerhard Knorz, FR Darmstadt, Prof. Dr. Jürgen Handke, Universität Marburg, PD Dr. Ulrich Schmitz, Universität Duisburg, Prof. Dr. Dietmar Rösner, TU Freiberg, Prof. Dr. Wolfgang Hoeppner, Universität Duisburg, Harald Elsen, Universität Bonn. Die Mitgliederversammlung nimmt die Kandidatenliste bei 3 Enthaltungen einstimmig an.

TOP 9: Arbeitsprogramm 1995/96

1996 wird eine Folgeveranstaltung der MORPHOL YMPICS mit Französisch als Haupttestsprache stattfinden. Veranstaltungsort wird ein Ort in der Nähe der deutsch-französischen Grenze sein. Die MORPHOL YMPICS wird voraussichtlich gemeinsam mit einer Abteilung des CNRS und der ATALA, der der deutsch-französischen Grenze sein. Die MORPHOLYMPICS wird voraussichtlich gemeinsam mit einer Abteilung des CNRS und der AT ALA, der französischen Schwesterorganisation der GLDV, durchgeführt werden. Eventuell wird auch ELSNET als Mitveranstalter auftreten. Die KONVENS 1996 wird von der Sektion Computerlinguistik der DGfS organisiert und in Bielefeld stattfinden. Das LDV-Forum 1/1995 wird voraussichtlich im Juni erscheinen. Der Vorstandsvorsitzende ruft die Mitglieder auf, sich mit wissenschaftlichen Beiträgen am LDV-Forum zu beteiligen. W. Lenders teilt mit, daß sich Herr Rösner von der TU Freiberg bereit erklärt hat, die Organisation der Herbstschule 1996 zu übernehmen.

TOP 10: Nächste Jahrestagungen

Die Jahrestagung 1997 der GLDV soll nach Möglichkeit in den neuen Bundesländern stattfinden. (Inzwischen hat sich Herr Heyer, Universität Leipzig, bereit erklärt, die Organisation und Durchführung der Jahrestagung 1997 zu übernehmen.)

TOP 11: Verschiedenes

Entfallt.

Uta Seewald
(Schriftführung)

Winfried Lenders
(Sitzungsleitung)

Mitglieder des Beirats der GLDV

(Stand: Dezember 1995)

Dr. H. Billing Projektträger
des BMBF für
Fachinformation
Dolivostr. 15
64293 Darmstadt
Tel.: (06151) 869732

Harald Elsen
Universität Bonn
Institut für Kommunikationsforschung und Phonetik
Poppelsdorfer Allee 47
53115 Bonn
Tel.: (0228) 735677
Email: hel@ikp.uni-bonn.de
WWW: http://cll.ikp.uni-bonn.de/-hel

Prof. Dr. Wolfgang Hoeppner
(Sprecher des Beirats)
Universität Duisburg
Fachbereich 3 - Computerlinguistik Postfach 10 15 03
47057 Duisburg
Tel.: (0203) 379-2006 oder: 2008
Email: hoeppner@unidui.uni-duisburg.de

Prof. Dr. Gerhard Knorz
Fachhochschule Darmstadt
Fachbereich luD
Haardtring 100
64295 Darmstadt
Tel.: (06151) 16-8499
Fax: (06151) 16-8980
Email: knorz@fh-darmstadt.de
WWW: http://www.iud.fh-darmstadt.de

Prof. Dr. Dietmar Rösner
Otto-von-Guericke-Universität Magdeburg
Fakultät für Informatik
Institut für Informations- und Kommunikationssysteme
Postfach 4120
39016 Magdeburg
Tel.: (0391) 67-18314
Fax: (0391) 67-12018
Email: roesner@iik.cs.uni-magdeburg.de
WWW: http://www-ai.cs.uni-magdeburg.de

Prof. Dr. Ulrich Schmitz
Universität GHS Essen
Fachbereich 3
Literatur- und Sprachwissenschaft
Universitätsstr. 12
45117 Essen
Tel.: (0201) 183-3422

Dr. Matthias Wermke
Bibliographisches Institut Mannheim
Dudenredaktion
Dudenstr.6
68167 Mannheim
Tel.: (0621) 3901420
Fax: (0621)3901430

ARBEITSKREISE DER GLDV

AK - Korpora

Korpus basierte Lexikographie

Arbeitskreis Korpora tagt am 8.6.95 an der Univ. Stuttgart, Institut für maschinelle Sprachverarbeitung - Computerlinguistik (IMS-CL)

Robert NEUMANN begrüßte in seiner Eigenschaft als Leiter des Arbeitskreises "Korpora" innerhalb der GLDV die Referenten und die Gäste und dankte dem Institut für maschinelle Sprachverarbeitung der Universität Stuttgart für die Ausrichtung des Treffens. Er stellte erfreuliche Kontinuität in der wissenschaftlichen Diskussion des Arbeitskreises fest, die sich in diesem sechsten Treffen unter seinem Vorsitz manifestiert. Es folgte die Begrüßung durch den Gastgeber Ulrich HEID, der in groben Zügen sein Institut vorstellte, das im Jahre 1987 als Forschungszentrum für zentrale Fragen der theoretischen Linguistik und der maschinellen Bearbeitung von natürlicher Sprache gegründet wurde. Es besteht aus den Abteilungen "Computerlinguistik" (Prof. Christian Rohrer), "Experimentelle Phonetik und Phonologie" (Prof. Grzegorz Dogil), "Logik und Sprachphilosophie" (Prof. Hans Kamp) und" Theoretische Computerlinguistik" (Prof. Mats Rooth).

Robert NEU MANN berichtete in seinem Vortrag "MECOLB (Multilingual Environment for Corpusbased Lexicon Building): Ein Schritt auf dem Weg zur korpusbasierten Lexikographie" über das EU-Projekt COSMAS-II, dessen Projektkoordinator er selbst ist, und das Konzept der virtuellen Korpora und ihrer Multidimensionalität. Im einzelnen beschrieb er die fachlichen, innovatorischen, statistischen, informatischen und linguistischen Dimensionen von COSMAS und die Möglichkeiten, die das erweiterte System für die Verifizierung und Falsifizierung von linguistischen Hypothesen bietet.

Cyril BELICA ("Statistische Analyse von Zeitstrukturen in Korpora") hat am Institut für deutsche Sprache im Zusammenhang mit dem Lexikographie-Projekt "Wendewortschatz" eine Methode zur Berechnung, Analyse und Auswertung zeitrelevanter statistischer Parameter von Texten entwickelt und implementiert, die er sehr anschaulich vorstellte. Auf einem konkret festgesetzten Signifikanzniveau werden sprachliche Einheiten isoliert, deren zeitliche Verteilung im Text einer durch den Sprachwissenschaftler angenommenen "natürlichen" Verteilung widerspricht. Mit diesen mathematisch-statistischen Aussagen über einen Sachverhalt von Zeichenketten im Text können sehr leicht bereits mehr oder weniger stark frequentierte, aber auch potentielle Neologismen in den Korpustexten - das Analysekorpus umfaßte 4 Millionen laufende Wörter - ermittelt werden. Es werden dadurch Evidenzen bei der synchronischen und diachronischen Sprachanalyse aufgezeigt; die Signifikanz solcher Auffälligkeiten kann somit besser beurteilt werden, und explizite Hinweise auf die Interpretation dieser Auffälligkeiten werden möglich.

Der Vortrag machte klar, daß das vorgestellte Verfahren eine Analyse der zeitlichen Entwicklung von Sprachphänomenen ermöglicht. Es kann für das automatische

Filter von Texten für Monitorkorpora, als Programm zur Überwachung externer Textquellen und zur automatischen Aktivierung der Aufzeichnungseinrichtungen von Nachrichtendiensten angewendet werden.

Ulrich HEID sprach über" Texterschließungswerkzeuge als Hilfsmittel für den korpusbasierten Aufbau von Wörterbüchern"; seine Vorbemerkungen betrafen den institutionellen Rahmen zum Projekt "Textkorpora und Erschließungswerkzeuge", die Softwareentwicklung und die Entwickler. Es herrsche Einigkeit über die Notwendigkeit und die Vorteile einer korpusbasierten Lexikographie, aber bisher seien nicht sehr viele Werkzeuge zur Unterstützung der korpuslexikographischen Arbeit entwickelt worden. Eine alternative Vorstellung zum einmal erstellten und nie veränderten Wörterbuchaufbaumodell ist die ständige Verfeinerung der lexikalischen Modellierung der inkrementellen Daten, was eine iterative und zunehmend verbesserte Korpusabfrage bedeutet. Ulrich HEID entwickelte die Arbeitsschritte für die linguistische Korpusanalyse und -annotation sowie die Abfrage mit Hilfe von Konkordanzwerkzeugen. Die zur Zeit verfügbaren Tools sind Tokenizing (Wortformen, Satzgrenzen), eine morphosyntaktische Analyse und Lemmatisierung, ein Part-of-Speech-Tagging, ein partielles Parsing zur Erkennung phrasenstruktureller Konstrukte. Als Anwendungsbeispiel führte der Referent ein Lexikonfragment für lexikalisierte und nichtlexikalisierte Funktionsverbgefüge vor.

Dr. Achim STEIN ("Französische und italienische Korpora") berichtete über die Erstellung der Korpora, ihre Ressourcen und Anwendungsbereiche. Zur Erstellung der Korpora gehören die Erarbeitung eines Grundformenlexikons mit Tags für die Flexionsklassen, die Expansion zum Vollformlexikon (maximales Tagset), die Aufbereitung des Textes und die morphologische Annotierung. Als Hauptprobleme nannte der Referent im Zusammenhang mit dem Französischen und Italienischen die Behandlung von Eigennamen und Abkürzungen, von Mehrwortlexemen und agglutinierenden italienischen Pronomina und problematische Worttrennungen.

Oliver WAUSCHKUHN ("Statistisches Tagging als Vorstufe zur partiellen syntaktischen Analyse deutscher Texte: Einfluß auf die Anzahl der Parse-Ergebnisse") ging der Frage nach, ob der Einsatz eines Taggers als Präprozessor der syntaktischen Analyse zur A-priori-Disambiguierung für Parse-Ergebnisse dienen kann, denn für die linguistische Erschließung von Textkorpora z.B. durch Lexikographen wird zusätzlich zu den Tags eine strukturelle (syntaktische) Annotation benötigt. WAUSCHKUHN untersuchte satzweise die partielle syntaktische Analyse eines kleinen Korpus (aus der "Stuttgarter Zeitung" 3776 Sätze) jeweils in getaggter und in morphosyntaktisch ambig annotierter Form und verglich quantitativ die Anzahl der gelieferten Ergebnisse zwischen getaggter und nicht getaggter Form des Textes Als Schlussfolgerung schlug er entweder eine Verbesserungsmöglichkeit des Taggers oder ein zusätzliches Anbringen von Annotationen bei den zu erwartenden Fehltags und die Festlegung einer klaren Schnittstelle zwischen der Korpusannotation und der syntaktischen Analyse (Grammatik) sowie eine Verbesserung der NP-Grammatikregeln vor.

Petra STEINER gab im Vortrag "Das Münster-Tagging-Projekt (MTP)" einen Überblick über aktuelle Arbeiten am deutschen Textkorpus an der Westfälischen Wilhelms- Universität Münster, über die Aquirierung und Bearbeitung der ca. 100 Millionen Token - vorwiegend aus der "Frankfurter Allgemeinen" und der" Zeit" der Jahre 1990 bis 1992 - und über die Ergebnisse des automatischen Taggings. Ein verhältnismäßig geringer, zahlenmäßig aber beachtlicher Anteil wird mit Hilfe von bereits vorliegenden Informationen automatisch getaggt und mit einem speziellen Editor intellektuell disambiguiert bzw. korrigiert. Der größte Teil des Korpus soll lediglich mit Hilfe von automatischen Verfahren annotiert werden. Das Tagset genügt folgenden Anforderungen: klare Struktur mit mnemotechnischer Qualität, Operationalisierbarkeit und Intersubjektivität, Vorhersagbarkeit, Klasseneinteilung, Genauigkeit, Kompatibilität und Adaptabilität.

Holger TREBBE erhellte in seinem Beitrag "Das Hidden-Markov-Modell und seine Entwicklungsmöglichkeiten" die programmtechnische Seite der automatischen Verfahren. Innerhalb des "Münster Tagging Project" existieren momentan zwei Tagger, ein Counting-Rate-Tagger und ein HMM-Tagger. Beide Modelle beruhen auf der Voraussetzung zweier gekoppelter stochastischer Prozesse, die die Wortformenfolge/ Ambiguitätsklassenfolge und die Wortartenfolge beschreiben. Die Ereignisse der Wortformenfolge sind nicht direkt voneinander abhängig, sondern nur über die Wortartenfolge. Dabei wird auf der Basis der Bigramme gearbeitet. Da der Counting-Rate-Tagger mit relativen Häufigkeiten arbeitet und der HMM- Tagger auf dem Maximum-Likelihood-Konzept beruht, läßt sich die Korrektheitsrate erhöhen, wenn als Initialwerte des HMM- Taggers die ausgezählten relativen Häufigkeiten benutzt werden. Als Weiterentwicklung soll das VQHMM implementiert werden, das zusätzlich zur obigen Abhängigkeitsstruktur noch die Abhängigkeit der Wortformenfolge auf Basis der Bigramme unterstellt. Als Zusatzinstrument wurde eine Smoothing-Funktion vorgestellt, die verhindert, daß zu schätzende Parameter den Wert Null annehmen. Weiterhin wurde eine Gütefunktion dargestellt, die es ermöglicht, die Effektivität von Tag-Algorithmen trotz unterschiedlicher Tagsets und Testkorpora zu beurteilen.

Dr. Folker CAROLI sprach über "Korpusgestützte Grammatikentwicklung im Projekt LS-GRAM (Large Scale Grammar Development)", ein Projekt des Institut der Gesellschaft zur Förderung der Angewandten Informationsforschung e.V. an der Universität des Saarlandes (IAI) und zeigte anhand von Beispielen, wie bestimmte Methoden der Korpusuntersuchung die Grammatikuntersuchungen unterstützen. Die Eigenschaften der Anwendungsorientiertheit, der Erweiterbarkeit und der Wiederverwendbarkeit gelten für die in beiden Bereichen behandelten sprachlichen Phänomene - wie beispielsweise die Behandlung von Appositionen und Parenthesen.

Zum Abschluß des Arbeitstages führte Oliver CHRIST die Datenbank und das Recherchesystem des Instituts für maschinelle Sprachverarbeitung vor.

Irmtraud Jüttner, Mannheim

AK - Hypermedia

Das Treffen des Arbeitskreises Hypermedia der GLDV am 18.9.1995 in Mannheim

Der Arbeitskreis Hypermedia fand sich am 18. September 1995 bei freundlichem Spätsommerwetter zu einem ersten, konstituierenden Treffen im Institut für deutsche Sprache in Mannheim ein. Das Treffen diente vornehmlich dazu, sich kennenzulernen, dabei gemeinsame Interessen und Themen zu erkunden und über künftige Aktivitäten zu beraten.

Bei der Vorstellungsrunde wurden folgende Projekte angesprochen: Das am Institut für deutsche Sprache durchgeführte Pilotprojekt Grundlagen eines grammatischen Informationssystems GRAMMIS, in dem grammatische und lexikalische Datenbestände des IDS als Hypermedia Anwendungen präsentiert und mit verschiedenen Benutzergruppen getestet werden (A. Storrer); das am GMD-Institut für Integrierte Publikations- und Kommunikationssysteme (IPSI) angesiedelte, EU-geförderte Projekt Editors Workbench, in dem im Rahmen des Aufbaus eines wissensbasierten Publikationssystems eine Hypermedia-Kunstenzyklopädie entsteht (W. Möhr); das EU-geförderte Projekt LOTOS Learning Tools, bei dem u. a. an der Universität Tübingen ein modular aufgebautes Programmsystem für das computergestützte Fremdsprachenlernen konzipiert und entwickelt wird, das ein Lerner-Lexikon, eine pädagogische Grammatik und Werkzeuge zur automatischen Fehlerdiagnose umfaßt *(K.* Krüger-Thielmann); das LINGUA-Projekt GRECOTERM, in dem die Firma ZERES multimediale Lehrmaterialien für deutsche und griechische Touristikfachkräfte entwickelt sowie ein ebenfalls bei ZERES angesiedeltes Projekt, in dem multimediale Produktbeschreibungen für exportorientierte Weinkooperativen im

Rioja für das WWW erstellt werden (L. Lemnitzer); den von der Gruppo DIMA in Turin entwickelten NL-Parser DIMACHECK, für den am IAI Lingware zum Deutschen (Morphologie, Lexikon, Grammatik) entwickelt wurde, der als Syntaxchecker in eine Textverarbeitungssystem oder in Systeme für die computerunterstützte Fremdsprachenvermittlung integriert und mit Hyperdokumenten zu Grammatik und Lexikon des, Deutschen verbunden werden kann (S. Rieder, U. Reuther); das am wissenschaftlichen Zentrum der IBM in Heidelberg angesiedelte Projekt COALA (Computer Aided Language Acquisition), in dem u. a. an einer elektronischen Grammatik für das Deutsche gearbeitet wird, die mit einem als Hypertext organisierten Grammatiktutorial und einem Hypertext- Wörterbuch vernetzt wird (B. Harriehausen-Mühlbauer) .

Trotz der Verschiedenheit der Anwendungen (Publikationssysteme, Informationssysteme, Systeme zum computergestützten Fremdsprachenlernen, Grammatikprüfsysteme) ergaben sich erfreulich viele gemeinsame Fragestellungen, die im Rahmen des Arbeitskreises weiter diskutiert werden sollen: Wie kann grammatisches und lexikalisches Wissen mit Hypermedia modelliert und vermittelt werden? Wie lassen sich multimediale Gestaltungsmittel sinnvoll einsetzen? Wie können" traditionelle" Texte in Hypermedia-Anwendungen umgestaltet werden?

Ein weiterer Interessensschwerpunkt betrifft die Verbindung von lexiko- und grammatikographischen Hypermedia-Anwendungen zu Grammatikprüfsystemen und Systemen zum computerunterstützten Fremdsprachenlernen. Daneben besteht auch ein großes Interesse an Informations- und Erfahrungsaustausch im Bereich der Werkzeuge, mit denen sich Hypermedia Applikationen erstellen lassen. Ein Anliegen des Arbeitskreises wird es deshalb sein, Informationen über einschlägige Tools, Projekte, Diskussionslisten, Workshops und Tagungen zu sammeln und weiterzugeben. Als Sammelstelle hat sich das Institut für deutsche Sprache angeboten, die auch die WWW-Seite des AK (http://www.ids-mannheim.de/grammis/ak.html) weiter pflegt und darüberhinaus eine sogenannte Informationsbörse mit WWW-Adressen zum Thema anbietet.

Zum Abschluß des Treffens wurde die aktuelle Version des GRAMMIS-Systems vorgeführt, die neben grammatischen Informationen zu den Wortarten und den kommunikativen Funktionen des Deutschen auch ein elektronisches Glossar grammatischer Termini anbietet und über Schnittstellen zu einer Datenbank der deutschen Funktionswörter und zu einer Valenzdatenbank verfügt.

Als nächste größere Aktivität ist ein Arbeitstreffen zum Thema "Hypermedia in Grammatikographie und Lexikographie" geplant, das am 21. und 22. März am Institut für deutsche Sprache in Mannheim stattfindet; an der Organisation sind Bettina Harriehausen-Mühlbauer, Wiebke Möhr und Angelika Storrer beteiligt. In diesem Workshop möchten wir uns - aus theoretischer und aus anwendungsbezogener Perspektive - mit den neuartigen Gestaltungsmöglichkeiten befassen, die Hypermedia in den Bereichen Lexikographie, Terminographie und Grammatikographie eröffnet. Dabei interessieren uns konzeptionelle, textlinguistische und informatische Aspekte ebenso wie konkrete Anwendungen, z.B. Grammatiken und Lexika für Sprachlernsysteme, WWW-Lexika, innovative elektronische Wörterbücher oder lexikalische Datenbanken. Weiterhin gibt es Beiträge zu Methoden und Werkzeugen, mit denen vorhandene Publikationen in Hypermedia-Anwendungen umgesetzt werden können, zu Fragen der Evaluierung und Benutzungsforschung in diesem Bereich, zur Konzeption von CALL-Systemen sowie zu Gesichtspunkten der Ergonomie und des Grafik-Designs. Aktuelle Informationen zum Workshop finden sich unter http://www.ids-rnannheirn.de/grarnrnis/work.htrnl.

Angelika Starrer, Heidelberg