

## ARBEITSKREISE DER GLDV

### AK - Studienführer CL

#### Aufruf zur Mitarbeit

Nachdem seit dem Erscheinen des letzten Studienführers Computerlinguistik bereits einige Zeit verstrichen ist und sich in der deutschen CL-Landschaft auch institutionelle Veränderungen ergeben haben, hat der Vorstand der GLDV auf der diesjährigen Jahrestagung in Regensburg beschlossen, den Arbeitskreis Studienführer zu reaktivieren und eine neue Auflage des Studienführers in Angriff zu nehmen.

Folgender Arbeitsplan ist vorgesehen:

#### *O. Definition eines geeigneten Formates für den Studienführer*

Da anders als in der vorangegangenen Auflage die neue Version auch in elektronischer Form zugänglich sein soll, erscheint es ratsam, die Informationen in SGML bzw. HTML aufzuarbeiten. Der Entwurf einer zu diesem Zweck geeigneten DTD ist in Arbeit.

#### *1. Erfassung vorhandenen Materials*

Wenigstens teilweise können die Daten der alten Auflage übernommen werden, müssen allerdings in das neue Format überführt werden.

#### *2. Ermittlung neuer Daten*

Per Rundschreiben an die Hochschulen bzw. durch eine Telephonaktion sollen die neuen/ geänderten Informationen zur Computerlinguistik ermittelt werden.

#### *3. Evaluierung der Informationen*

Es scheint erforderlich, nach Abschluß

der Datensammlung eine gewisse Evaluierung durchzuführen, um dem Status und der tatsächlichen Bedeutung der ermittelten Veranstaltungen und Studiengänge gerecht zu werden (z.B. Studienschwerpunkt innerhalb eines Faches vs. eigener Studiengang; seltene Spezialvorlesung vs. regelmäßiger Vorlesungszyklus etc.).

#### *4. Aufarbeitung als Informationsangebot im WWW*

Um auf möglichst einfache Weise weiteres Feedback erhalten zu können und gleichzeitig die Information einem möglichst breiten Interessentenkreis zugänglich zu machen, werden alle Daten sowie die Datenstruktur für Neueinträge im WorldWideWeb verfügbar gemacht. Einstiegspunkt ist zunächst der WWW-Server des Instituts für Informatik der Universität Leipzig (<http://www.informatik.unileipzig.de>), weitere werden evtl. folgen. Je nach verfügbarer Zeit und bestehendem "Benutzerinteresse" sind natürlich unterschiedlich detaillierte (und elegante) Lösungen denkbar Anregungen sind jederzeit willkommen!

#### *5. Drucklegung einer neuen Auflage* Sobald die obigen Punkte abgearbeitet sind, kann die Drucklegung der neuen Auflage erfolgen.

Alle an der Erarbeitung und Ausgestaltung des neuen Studienführers Interessierten sind herzlich eingeladen (und ermuntert), am Arbeitskreis teilzunehmen - geplant ist ein Treffen möglichst noch vor der

Sommerpause, genauer Termin und Ort stehen allerdings noch nicht fest. Nähere Informationen finden sich demnächst unter der genannten WWW-Adresse in Leipzig.

Die Kontaktadresse für den Arbeitskreis ist (bevorzugt e-mail):

Dr. Christian Wolff  
 Institut für Informatik  
 Abt. Automatische  
 Sprachverarbeitung Universität  
 Leipzig Augustusplatz 10-11  
 04109 Leipzig  
 Tel. (0341) 97-32249 (0. -32230)  
 Fax (0341) 97-32209  
 e-mail:  
 wolff@inform1.informatik.uni-  
 leipzig.de

## AK - Korpora

### Treffen des Arbeitskreises KORPORA am 21. Februar 1995 im Institut für deutsche Sprache (IDS), Mannheim

Am 21. Februar 1995 tagte der Arbeitskreis "Korpora" der Gesellschaft für Linguistische Datenverarbeitung im Institut für deutsche Sprache, Mannheim, und beschäftigte sich mit der Thematik "Statistischer Zugang zu Korpora". Es nahmen teil:

Prof. Dr. Ursula Klenk (Universität Göttingen),  
 Prof. Dr. Winfried Lenders (Universität Bonn),  
 Kathrin Giesecking (Universität Trier),  
 Alexander Mehler (Universität Trier),  
 Markus Schulze (Universität Erlangen),  
 Gerald Schüller (Universität Erlangen),  
 Jochen Leidner (Universität Erlangen),  
 Helmut Schmid (Institut für maschinelle Sprachverarbeitung,

Universität Stuttgart),  
 Stefan Langer (Centrum für Informations- und Sprachverarbeitung, Universität München),  
 Robert Naumann, Dons al-Waldi und Dr. Irmtraud Jüttner, jeweils IDS, Mannheim

Herr Robert Neumann (Leiter des Arbeitskreises) eröffnete die eintägige Zusammenkunft und umriß das Ziel der aktuellen Diskussion des Arbeitskreises, verschiedene Möglichkeiten statistischer Analysen vorzustellen. Die Arbeiten aus dem Bereich Computerlinguistik/LDV der Universität Trier wurden von Frau Kathrin Giesecking und Herrn Alexander Mehler vorgestellt.

Frau Giesecking sprach zum Thema "Empirische Überprüfung eines synergetischen Modells der Lexik anhand englischer Korpora", des Modells von Reinhard Köhler, das für die Betrachtung der Lexik die folgenden Systemgrößen "Frequenz", "Länge", "Polylexie", "Phonemanzahl" und "Lexikongröße" entwickelt hat.

Die synergetische Linguistik betrachtet die Sprache als ein sich selbst organisierendes System, das Mechanismen für eine Anpassung an die Systemumgebung entwickelt und gewisse Ansprüche an die Sprache - auch divergierende Systemansprüche - stellt. Und so sorgen Selbstorganisationsmechanismen für eine Anpassung an den Sollzustand, beispielsweise Mutations- und Selektionskriterien beim Silbenverschlucken. Neben dem Anpassungsbedürfnis erläuterte Frau Giesecking die Stabilitäts-, Kodierungs-, Spezifikations-, Anwendungs- und Ökonomiebedürfnisse, das Bedürfnis nach Sicherheit der Informationsübertragung und die speziellen Bedürfnisse nach Minimierung des Kodierungsaufwands. Die empirische Überprüfung des Basissystems von Köhler zeigt eine gerichtete Abhängigkeit zwischen der Frequenz, der Länge, der Polylexie und Polytextie. Zwischen Frequenz und Silbenlänge gibt es große Zusammenhänge; so wird bei häufigem Gebrauch des mehrsilbigen "Universität" ein Wort mit der Silbenlänge 2, nämlich "Uni". Herr Alexander Mehler (Universität Trier) beleuchtete in

seinem Vortrag "Fuzzy-Textbedeutungen" Probleme der Textbedeutungsrepräsentation und erarbeitete die Hypothese, daß die Zweistufigkeit der Bedeutungskonstitution, welcher die syntagmatischen und paradigmatischen Restriktionen der Sprachelemente untereinander entsprechen, auch auf der Ebene von Texten Geltung besitzt. Diese Zweistufigkeit führt auf der Textebene zur Konstitution der konnotativen Textbedeutung, deren Konstituenten auf der Wortebene durch die syntagmatischen und paradigmatischen Relationen der Wörter bestimmt werden.

Der von Herrn Mehler vorgeschlagene Repräsentationsformalismus für Textbedeutungen kann in einem weiteren Schritt zur Lösung des Problems der automatischen Textkonversion im Rahmen des Hypertextes herangezogen werden. Ein alternatives Verfahren zur bisherigen intellektuellen Textkonversion sieht er in einem automatischen Aufbau der Hypertextbasis durch textoberflächenorientierte Informationen, womit zumeist die Notwendigkeit einer Nachedition oder Überprüfung der resultierenden Hypertextstruktur durch den Informationswissenschaftler verbunden ist. Weil der Rekurs auf rein formale Textmerkmale - etwa die Untergliederung der Texte in Abschnitte - zu einer weitgehenden Ausklammerung des Textbedeutungsaspektes führt, schlägt Herr Mehler ein Verfahren zur automatischen Textkonversion vor, das primär auf Repräsentationen der Bedeutungen der zu verknüpfenden Wörter und Texte basiert und somit den Bedeutungsaspekt in den Mittelpunkt der Betrachtung rückt.

Herr Helmut Schmid (Universität Stuttgart) berichtete in seinem Referat über "Aktivitäten am Institut für Maschinelle Sprachverarbeitung (IMS) im Bereich des Taggens deutscher Texte". Unter Tag ging subsumiert er alle Annotationen von Textkorpora mit syntaktischen, morphologischen, semantischen und anderen Informationen. Beim Wortarten- Tagging treten die bekannten Probleme mit den syntaktischen ambigen Wortformen auf, und die einzige Lösung wäre eine Desambiguierung durch den Kontext - mit

den Möglichkeiten durch Markov-Ketten, durch neuronale Netze, durch Regelbasierung und durch einen Parser. Er diskutierte Vor- und Nachteile verschiedener Trainingsmodelle (Baum-Welch-Training, Viterbi-Training, Xerox-Tagger).

Herr Stefan Langer (Centrum für Informations- und Sprachverarbeitung, Universität München) sprach zu "Statistischen Untersuchungen der Semantik von Nominalkomposita im Rahmen der semantischen Kodierung der CISLEX". Das Kernlexikon enthält 38 000 einfache Nomina für die statistische Untersuchung der Semantik. Bei ersten Versuchen einer Einteilung ließen sich bisher ungefähr 250 semantische Klassen unterscheiden mit einer Grobgliederung in Lebewesen, andere Artefakte, Orts- und Zeitbezeichnungen. Herr Langer demonstrierte an Beispielen die statistische Analyse von Selektionspräferenzen in den Nominalkomposita.

Herr Neumann faßte den Vortrag "Statistische Analyse von Zeitstrukturen in Korpora" von Herrn Cyril Belica (IDS, Mannheim) kurz zusammen und beantwortete Fragen. Herr Belica, der wegen Krankheit kurzfristig seinen Vortrag absagen mußte, hat eine Methode zur Berechnung, Analyse und Auswertung zeit relevanter Parameter von Sprachphänomenen in Textkorpora, die als Indikatoren des Terminus "Neologismus" verwendet werden, entwickelt. Diese mathematisch-statistische Analyse zeitlicher Verteilung von Häufigkeiten verschiedener Textmerkmale isoliert solche Sprachelemente, deren zeitliche Verteilung auf dem festgesetzten statistischen Signifikanzniveau bestimmte Auffälligkeiten aufweist. Herr Neumann erklärte die programmtechnische Implementierung der Methode, gab Beispiele für die statistische Berechnung und Findung der sogenannten Kandidaten für einen Neologismus im deutschen Wortschatz und für eine mögliche Kollokationsanalyse.

Frau Doris al-Wadi führte eine COSMAS-Präsentation vor und zeigte an relevanten Sprachbeispielen aus den Textkorpora des IDS - insbesondere am "Wendekorpus", das in sechs zeitlichen Phasen abfragbar ist - die statistischen Sor-

tierungen und Gruppenbildungen bei unterschiedlichem Signifikanzniveau sowie bei den veränderbaren Parametern der Granularität (Clusterbildung) und Klassenzuordnung; zur besseren Illustration dieser Methode gab sie den Teilnehmern eine vorbereitete Tischvorlage zur Kollokationsanalyse in die Hand.

Das nächste Treffen des Arbeitskreises "Korpora" ist für den 8. Juni 1995 an der Universität Stuttgart, im Institut für maschinelle Sprachverarbeitung - Computerlinguistik, geplant, wo über "Korpusbasierte Lexikographie" verhandelt werden soll.

*Irmtraud Jüttner, Mannheim*

#### **AK - Fuzzy Linguistik**

##### **Ankündigung des neuen Arbeitskreises und Aufruf an Interessenten zur Mitarbeit**

Während der beiden letzten GLDVJahrestagungen (Kiel 1993, Regensburg 1995) hat es jeweils eine Sektion *Fuzzy Linguistik* gegeben, deren Vorträge auf sehr erfreuliche Resonanz stießen. Da auch mein Vorschlag ein positives Echo fand, diesem Interesse ein zwischen den Jahrestagungen vorhandenes Forum zu bieten, möchte ich hiermit die - schon in Regensburg angekündigte - Einrichtung des neuen Arbeitskreises *Fuzzy Linguistik* in der GLDV bekanntgeben.

*Fuzzy* linguistische Modellbildungen erfassen zunehmend mehr und weitere Bereiche. Anfänglich nur im Rahmen einer traditionellen Semantiktheorie zur Analyse, Repräsentation und Verarbeitung natürlichsprachlicher Unschärfe eingesetzt, bilden heute die methodisch wie methodologisch (weiter-)entwickelten *fuzzy* Theorien ein neues Paradigma, das fast alle logischen und mathematischen, insbesondere aber die empirisch-numerischen Methoden und Verfahren beeinflusst, welche für die Modellbildungen in der Linguistik, der Computerlinguistik und der quantitativen Linguistik seit langem eingesetzt werden. Dieser

Einfluß scheint derzeit noch verstärkt zu werden durch eine merkliche Hinwendung linguistischer Forschung zu performativen Phänomenen des Sprachgebrauchs und deren funktionaler Erklärung in kognitivkommunikativen Zusammenhängen.

Welche Konsequenzen eine zunehmende Übertragung von *fuzzy* Modellierungstechniken für unsere Disziplin (wissenschaftstheoretisch wie praktisch) haben wird, ist wohlfundiert kaum abschätzbar und kann nur vage umrissen werden. Denkbar ist aber immerhin, daß ihr vermehrter Einsatz zu einer schrittweisen Revision des (wissensbasiert-kognitiven) Erkenntnisinteresses der Linguistik (zugunsten eines lernensbasiert-semiotischen) beitragen wird angesichts der schon jetzt beobachtbaren Erweiterungen, welche die (kompetenztheoretische) Bestimmung der linguistischen Forschungsgegenstände (in Richtung auf performative Sprachdaten) erfährt, und angesichts der *fuzzy* Modellierungen, deren integrative, kontinuierliche und dynamische Repräsentationen die (auf inkompatiblen Zwischenrepräsentationen operierenden) Modellbildungen der bisherigen Untersuchungsmethoden zunehmend ergänzen und sogar ablösen.

Die im Schnittbereich von kognitiver Linguistik und kognitiver Sprachverarbeitung mit Computerlinguistik und quantitativer Linguistik sich konstituierenden neuen (oder auch nur veränderten) *E-Interessen*, *F-Gegenstände* und *U-Methoden* könnten darüber hinaus eine Linguistik etablieren, die bei der Erforschung natürlicher und der Entwicklung künstlicher informationsverarbeitender Systeme auch anwendungsrelevante Problemlösungen zu bieten vermöchte. Beides, *fuzzy* Modellierungen als Vorbereiter einer semiotisch-linguistischen Sprachtheorie und als Basis einer linguistisch-simulativen Ingenieurwissenschaft markieren die Pole linguistischer Forschung, Entwicklung und Anwendung, zwischen denen sich die Aktivitäten des Arbeitskreises *Fuzzy Linguistik* bewegen könnten.

Interessenten und potentielle Mitglieder des AK mögen sich bitte per Email, Fax

oder auch brieflich mit mir in Verbindung setzen. Weil die Arbeitskreise der GLDV allen, insbesondere auch den jüngeren Mitgliedern offenstehen, bitte ich gleichzeitig auch um informative Angaben zu Hintergrund und Motivation für die Meldung zum AK Fuzzy Linguistik. Je nach Anzahl und Interessenlage bei diesen Anmeldungen möchte ich nach Möglichkeit noch während des laufenden Semesters (spätestens zu Beginn des kommenden Wintersemesters) ein erstes Arbeitstreffen hier in Trier organisieren. Es soll neben dem wechselseitigen Kennenlernen und der Vorstellung der jeweils eigenen Forschungsinteressen und Arbeitsansätzen vornehmlich dazu dienen, eventuell konvergierende Bereiche ausfindig zu machen und anhand erkennbarer (möglicher) Schwerpunkte über die künftige Arbeit und das Programm des AK zu diskutieren.

*Burghard Rieger*

Linguistische Datenverarbeitung  
Lehrstuhl für Computerlinguistik FB 11:  
Sprach- und Literaturwissenschaften,  
Universität Trier D-54286 Trier  
*Email:* fulinak~ldv35.Uni-Trier.de Fax:  
+49-651-201-3946  
*Tel:* +49-651-201-2270

## AK - Hypermedia

### Der Sommer kommt, ein Arbeitskreis erwacht: Der Arbeitskreis Hypermedia

#### *Warum ein Arbeitskreis Hypermedia?*

Im Zuge des Aufbruchs in die sog. Informationsgesellschaft arbeiten verschiedene Fachrichtungen gemeinsam an der Frage, wie elektronische Dokumente in neuartiger Weise organisiert und präsentiert werden können. Auch die Computerlinguistik kann zur Entwicklung qualitativ hochwertiger Hypermediasysteme und -anwendungen beitragen, z.B. durch Textparser zur (semi)automatischen Konversion konventioneller Texte in Hyperdokumente

oder durch intelligente, natürlichsprachliche Navigationshilfen. "Einfache" LDV-Verfahren wie morphologische Analyse oder Lemmatisierung sind wertvoll, wenn beispielsweise Termini in Hypertexten automatisch mit Glossaren oder Wörterbüchern verknüpft werden. Darüberhinaus bietet Hypermedia in verschiedenen Bereichen der angewandten Computerlinguistik völlig neuartige Möglichkeiten, Arbeitsabläufe zu organisieren und die dabei entstehenden Ergebnisse zu präsentieren.

Anlaß genug, das Arbeitskreisgeschehen der GLDV um einen Arbeitskreis zu bereichern, der sich speziell mit den computerlinguistischen Aspekten der Hypermedia-Technologie auseinandersetzt.

*Wer sollte sich angesprochen fühlen?*

>Entwicklerinnen und Entwickler von Hypermedia-Anwendungen in Bereichen der angewandten Computerlinguistik, z.B. computergestützte Lexikographie, Terminographie und Grammatikographie oder computergestützter Fremdsprachenunterricht .

> Personen, die Hypermediasysteme entwickeln und Interesse an der Integration computerlinguistischer Werkzeuge und/oder wissensbasierter natürlichsprachlicher Navigationshilfen haben.

> Personen, die mit der Konversion von Texten in Hyperdokumente bzw. der (Re ) Konversion von Hypertexten in Texte befaßt sind.

*Womit wird sich der AK Hypermedia befassen?*

Was im einzelnen getan wird (werden kann), hängt von der Anzahl der AK-Mitglieder und deren Interessen ab. Denkbare Aktivitäten sind:

> Bestandaufnahme existierender und geplanter Hypermedia- Anwendungen im Bereich der angewandten Computerlinguistik

> Austausch von Erfahrungen, Tools, Programmen

- > Hypermediasysteme als neues Anwendungsfeld für computerlinguistische Werkzeuge und Methoden: Ideensammlung, Befragung von Hypermedia-Entwickler(innen)
  
- > Organisation themenspezifischer Workshops

Wer am Arbeitskreis mitarbeiten möchte, möge sich möglichst bald bei der unten angebenen Kontaktadresse melden. Für Mitte September ist ein erstes, konstituierendes Treffen in Mannheim geplant. Näheres ist ab Juli unter <http://www.ids-mannheim.de/projekt/grammis/ak-hypermedia.html> zu erfahren.

**Kontakt:**

Dr. Angelika Storrer Institut  
für deutsche Sprache Abteilung  
Grammatik  
PO 101621  
D-68016 Mannheim  
Tel. 0621-1581207  
Fax: 0621-1581200  
e-mail: [storrer@ids-mannheim.de](mailto:storrer@ids-mannheim.de)