

Ein topikbasiertes Retrievalsystem in Hypertext

Heinz J. Weber, Constantin Thiopoulos
e-mail: weber@utrurt.uucp.de

1 Einleitung

Retrievalsysteme, die auf booleschem Retrieval basieren, sind - im Hinblick auf die exponentiell wachsenden Mengen von gespeicherten Informationen - nicht effizient, da sie nur in der Lage sind, durch eine boolesche Kombination von keywords eine Menge von gespeicherten Dokumenten zu bestimmen. Der Benutzer muß dann aus diesen Dokumenten die Stellen, die ihn interessieren, manuell herausfinden. Außerdem muß er die "richtige" Abfrage stellen.

Hier wird dagegen ein Retrievalsystem vorgestellt, das:

- Die gespeicherten Dokumente automatisch in Topiks aufteilt, eine entsprechende Topikhierarchie erstellt und diese als Hypertextstruktur generiert. Dadurch ist es möglich, daß der Benutzer nur einen Topik eines Dokumentes zu lesen hat, in dem ein keyword vorkommt. Falls er daraus schließt, daß auch andere Teile des Dokumentes für ihn von Interesse sind, kann er durch Aktivieren der entsprechenden Hypertextverbindungen zu diesen gelangen, sonst geht er zu dem entsprechenden Topik des nächsten Dokuments.
- Das Retrieval genauer macht durch das Vorschalten eines Kollokationensfilters von Lexemen, die für den Benutzer interessant sind.
- Mit Hilfe eines Hypertextsystems das Retrieval auf das Aktivieren von graphischen Repräsentationen und Navigieren zurückführt, so daß der Benutzer die für ihn relevante Information bestimmen kann, ohne eine Abfrage explizit formulieren zu müssen.

2 Topikhierarchien

Texte werden oft nicht als isolierte Informationseinheiten verarbeitet, sondern können eine große Vielfalt zwischentextlicher Beziehungen aufweisen, die den Rezipienten in die Lage versetzen, durch Überschreiten der

Einzeltext-Grenzen Informationen zu sammeln und zu integrieren. Dies trifft u. a. auf wissenschaftliche Artikel zu, insbesondere jedoch auf Nachrichten. Hier besteht bei der Sammlung und Integration von Informationen die Möglichkeit, sowohl Texte der selben Tagesausgabe als auch Texte von zeitlich verschiedenen Ausgaben miteinander zu verbinden. Generell hängt es vom jeweiligen Leser ab, welche Texte in welcher Weise miteinander verbunden werden. Aber auch in den Texten selbst können Markierungen und Wegweiser vorkommen, die zur Orientierung dienen. Häufig verbinden diese Verweise Texte miteinander, in denen das selbe oder ein verwandtes Thema (im folgenden als "Topik" bezeichnet) behandelt wird. In Nachrichten dürfte dieser Typ von Text-Text-Relationen der wichtigste sein. Verbindungen zwischen Informationseinheiten führen zu einer Hypertext-Struktur, einer nicht-linearen Anordnung von Texten, die von der sequentiellen Präsentation (wie sie z.B. bei Nachrichten üblich ist) abweicht. [6]

Texte in nicht-linearer Anordnung werden von Hypertext-Systemen verwaltet. Die zwischen den Texten bestehenden logischen Verbindungen müssen entweder intellektuell/manuell hergestellt oder maschinell ermittelt werden. Bei maschineller Überführung von vorhandenen Texten in ein Hypertext-System muß eine Sprachverarbeitungs-komponente vorausgesetzt werden, d.h. ein Text-Parser und ein Link-Generator.

Der Beitrag stellt ein experimentelles Hypertext-System vor, in dem Einzeltexte des Typs Nachrichten nach thematischen Gesichtspunkten maschinell vernetzt werden sollen. Im folgenden soll eine Topik-Analyse skizziert werden, der jeder Einzeltext unterzogen wird. Ergebnis der Topik-Analyse ist die Zerlegung eines Textes in etikettierte Textsegmente und deren Anordnung in einer Topik-Hierarchie.

Für den Aufbau der Topik-Hierarchie soll das Vorgehen an einem kleinen Textsample demonstriert werden. Die drei Texte sind semi-authentisch: um Ablenkung zu vermeiden, werden fiktive Eigennamen verwendet.

Beispiel 1

T1, S1

Unter der Leitung von H. Schulze traf eine Delegation der FCS-Fraktion in Gladz ein.

T1, S2

Dort begannen Gespräche mit Delegierten des Ostsylianischen Parlaments...

T2, S1

Eine Delegation der FCS-Fraktion besuchte das Ostsylianische Parlament in Gladz.

T2, S2

In einer Rede gab H. Schulze, Vorsitzender der FCS-Fraktion, der Hoffnung Ausdruck, daß die Grenze zwischen beiden sylianischen Staaten eine Grenze ohne Mauern und Stacheldraht werden wird.

T3, S1

Mitglieder der FCS-Fraktion besuchten das Ostsylianische Parlament in Gladz.

T3, S2

Es ist das erste offizielle Treffen von Parlamentariern beider sylianischer Staaten.

T3, S3

In der Zwischenzeit äußerte M. Meyer, Geschäftsführer der CFS, Zweifel bezüglich Kontakten dieser Art.

T3, S4

Der Besuch einer westsylianischen Delegation könnte Statusfragen berühren, die nicht zu leicht genommen werden sollten.

Es wird nun versucht, jeden Text in kohäsive Segmente zu zerlegen; das geschieht entweder durch Zusammenfassung von Sätzen zu größeren Informationseinheiten oder durch Markierung von "Schnitten" zwischen Sätzen, die dann als Segmentgrenzen betrachtet werden. Grundlage der Zusammenfassung sind das Vorkommen von Kohäsionsmitteln wie Konnektive oder Anaphora am Satzanfang (z.B. "dort" in T1: S2), die Rekurrenz von Lexemen und Eigennamen, sowie Indikatoren für einbettende und eingebettete Strukturen (z.B. Sprechhandlungsverben in Verbindung mit dem Konjunktiv bei Redewiedergabe; vgl. T3: S3 & S4). "Rekurrenz" ist vorerst in einem engen Sinn definiert als Wiederauftreten des selben Lexems oder Namens

(z.B. "Fraktion" in T2: S1 & S2) oder als Vorkommen von Mitgliedern der selben Wortfamilie in aufeinanderfolgenden Sätzen (z.B. "Delegation" - "Delegierte" in T1: S1 & S2). Abwesenheit dieser Kohäsionsmittel führt zu einem "Schnitt" zwischen Sätzen. Für das Textsample ergibt sich daraus folgende Gliederung.

Beispiel 2

T1 Segment 1

S1: ... eine Delegation ... S2:

Dort... Delegierten...

T2 Segment 1

S1: ... der FCS-Fraktion ... S2:

... der FCS-Fraktion ...

T3 Segment 1

S1: ... das ... Parlament ... S2:

... Parlamentariern ... Segment

2

S3: H' äußerte ... Zweifel ... S4:

,.. könnte ... sollten.

Es ist auch möglich, daß ein Text nach diesem Analyseschritt aus einem einzigen Segment besteht oder daß er so viele Segmente hat wie Sätze. Im Durchschnitt weisen die Texte unseres Testcorpus (300 Nachrichtentexte) vier Segmente auf, ein Segment besteht aus durchschnittlich drei Sätzen. Hinsichtlich der Ermittlung von Textsegmenten ist klar, daß es eine Reihe weiterer Indikatoren gibt, sowohl für Kohäsion als auch für Diskontinuität (z.B. Vorkommen von Synonymen wie 'besuchen' - 'Treffen' - 'Kontakte' in T3 oder die Abfolge von unbestimmtem und bestimmtem Artikel).[9] Jedoch beim derzeitigen Stand des Experiments sollen nur eine geringe Anzahl robuster und verlässlicher Informationen herangezogen werden.

Die skizzierte Textzerlegung liefert eine Text-Interpunktion, die als Grundlage für die weiteren Schritte dient, die Binnen-Analyse der Textsegmente und die Konstruktion einer Topik-Hierarchie.

Die Binnen--Analyse resultiert in einer Repräsentation hypothetischer Topiks und einer Etikettierung der Segmente, die zum Aufbau einer Topik-Hierarchie führt. Folgende Segment-Etiketten werden unterschieden: Haupttopik, Subtopik, Redewiedergabe, Annotation. Obwohl sie aufgrund verhältnismäßig einfacher struktureller

Kriterien ermittelt werden können (was nicht heißt, daß diese Aufgabe trivial ist, denn die entsprechenden Segmente können sich über mehrere Sätze erstrecken), lassen sich die Segment-Etiketten auf einer abstrakteren Beschreibungsebene interpretieren, z.B. als Indikatoren für spezifische Textualitätsmodi (wie Narration, Argumentation oder Deskription) oder für bestimmte Text-Instanzen (wie Autor oder Figur). Haupt- und Subtopik können als Vertextungen im Modus "Narration" angesehen und der primären Text-Instanz "Autor" zugeordnet werden. Narrationsindikatoren für Nachrichten sind Präteritum und dritte Person bei den Verben. Die Unterscheidung zwischen Haupt- und Subtopik beruht auf mehreren Faktoren. Ein Subtopik

- ist durch Adverbiale oder Konjunktionen eingeleitet (vgl. "in der Zwischenzeit" in T3),
- kann im Tempus wechseln (z.B. von Imperfekt zu Plusquamperfekt),
- führt neue Referenten ein (vgl. 'Meyer', 'CFS' in T3).
- steht meist in nicht-initialer Textposition,

Für eine Entscheidung zugunsten der Etikettierung als Subtopik genügen zur Zeit zwei der vier Faktoren. Die Bezeichnungen Haupt- und Subtopik sind nicht mißzuverstehen als Abstufungen hinsichtlich der inhaltlichen Relevanz eines Textsegments, sie bezeichnen nur die Position in einem syntaktischen Gefüge - analog zur Satz-Hierarchie "Hauptsatz" und "Nebensatz", die ja auch nicht als Gewichtung von Information anzusehen ist.

Für T3 aus dem Textsample ergibt sich folgende Darstellung:

Beispiel 3

T1 Segment 1
Haupttopik I: Mitglieder der FCS-Fraktion besuchten das Ostsylvanische Parlament in Gladz. ...

Segment 2

Subtopik I: In der Zwischenzeit äußerte M. Meyer, Geschäftsführer der CFS, Zweifel bezüglich Kontakten dieser Art. ...

Annotationen können als Vertextungen im Modus "Kommentierung" interpretiert werden. Indikator ist der Tempuswechsel vom Präteritum ins Präsens. Die entsprechende Text-Instanz für Kommentierung ist der "Autor".

Beispiel 4

T3 Segment 1

Haupttopik I: Mitglieder der FCS-Fraktion besuchten das Ostsylvanische Parlament in Gladz.

Annotation J: Es ist das erste offizielle Treffen von Parlamentariern beider sylvanischer Staaten.

Redewiedergabe ist einer untergeordneten, nicht-autonomen Textinstanz zuzuordnen (z.B. Sachverhaltsbeteiligte) und ist im Modus "Zitierung" vertextet. Zitierungsindikatoren sind Ausdrücke für Sprechhandlungen (z.B. "...äußerte... Zweifel") in Verbindung mit Konjunktionen, Konjunktiv oder Anführungszeichen. Segmente, die als Redewiedergabe etikettiert werden, sind wiederum als "Text" zu behandeln und in "Haupttopik", "Subtopik" etc. zu unterteilen, wobei die Möglichkeit von Rekursion via "Redewiedergabe" besteht. [10] [16]

Beispiel 5

T3 Segment 1

Haupttopik I: Mitglieder der FCS-Fraktion besuchten das Ostsylvanische Parlament in Gladz.

Annotation I: Es ist das erste offizielle Treffen von Parlamentariern beider sylvanischer Staaten.

Segment 2

Subtopik I: In der Zwischenzeit äußerte M. Meyer, Geschäftsführer der CFS, Zweifel bezüglich Kontakten dieser Art.

Redewiederg.: Haupttopik: Der Besuch einer westsylvanischen Delegation könnte Statusfragen berühren, die nicht zu leicht genommen werden sollten.

Die Identifikation und Isolation von Redewiedergabe in Texten ist ein sehr wichtiger Punkt beim Parsen narrativer Texttypen, wie Nachricht oder Bericht. Ganz abgesehen von dem Umstand, daß die schemafundierte (sogenannte wissensbasierte) Text-Analyse hier auf Schwierigkeiten stoßen muß, wenn beim "Auffüllen" von Schemaleerstellen die Grenze zwischen Sachverhalten innerhalb

und außerhalb von Redewiedergaben überschritten werden, auch im Kontext von Information Retrieval, Wissensextraktion [2] und maschineller Übersetzung [15] ist die Unterscheidung zwischen "originären" und "zitierten" Textpassagen zu beachten.

Im Anschluß an Textzerlegung, Binnenanalyse und Etikettierung der Segmente wird eine Repräsentation der in den etikettierten Textteilen residenten hypothetischen Topiks vorgenommen. Das Format ist zur Zeit noch primitiv: Es handelt sich um ein "Topik-Profil", eine Liste des jeweils vorkommenden Vokabulars - reduziert um die Strukturwörter. So ergibt sich für das Segment Subtopik (Redewiedergabe (Haupttopik)) in T3 folgendes Topik-Profil.

Beispiel 6

T3

Subtopik ...

Redewiedergabe:

Haupttopik: Besuch - westsylvanisch - Delegation - Statusfragen - berühren - leicht - nehmen

Natürlich kann ein solches "Topik-Profil" nicht als Repräsentation der Bedeutung eines Textteils angesehen werden wie eine "Makroproposition" etwa. Es handelt sich lediglich um die Oberfläche eines Konstrukts genannt "Topik", das Ergebnis der bisher genannten Analyseschritte ist. In Verbindung mit Text-Retrieval befindet sich diese Ebene der Topik-Beschreibung in einer Mittellage zwischen syntax-fundierte und schema-fundierte Strategien. Sie geht hinaus über die Betrachtung der Zusammengehörigkeit von Termen im Rahmen eines Syntagmas (z.B. Nominalgruppen, Präpositionalkonstruktionen) und sie verlangt nicht die Einbeziehung des ganzen Text-Vokabulars. Auch einige charakteristische Schwierigkeiten schema-fundierter Topik-Analysen lassen sich damit vermeiden. Diese treten dann zutage, wenn das Textmaterial den konzeptionellen Vorgaben nicht entspricht: Texte sind nicht immer stereotypen Inhalts und in schlichter Manier vertextet (wie die berühmten Schankschen Restaurant-Texte, um nur ein Beispiel zu nennen). Das gilt auch für Nachrichten. Im Hinblick auf die Nichtvorhersagbarkeit von Nachrichtenthemen und die vielfältigen Möglichkeiten der Vertextung ist es ratsamer, sich nicht allein auf vorfabrizierte Topik-Schemata zu verlassen, sondern genuin textliche Strukturelemente vorrangig zu verwenden.

Die Topik-Analyse schließt ab mit dem Aufbau eines Textbaums, einer Topik-Hierarchie. Zugrunde liegen dabei folgende Struktur-Regeln:

(1) $TEXT \rightarrow$
 $Haupttopik(&Subtopik*)(\&Annotation*)$

(2) $Haupt - /Subtopik \rightarrow$
 $Topik(\&Annotation)(\&TEXT)$

Nachfolgender Textbaum repräsentiert die Topik-Hierarchie in T3.

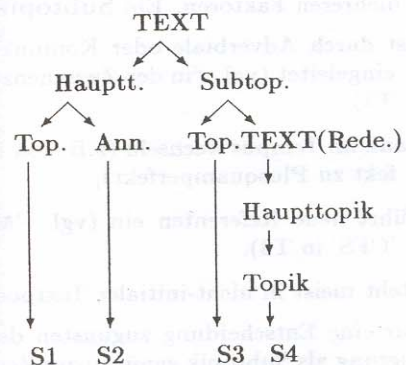


Abbildung 1: Topikhierarchie T3

3 Der Kollokationsgenerator

Durch die Generierung eines gewichteten Netzes, das die gegenseitige Abhängigkeit von Lexemen aufgrund von deren Gebrauch in den Texten repräsentiert, hat man ein Filter zu den Dokumenten zur Verfügung, das erlaubt:

- Das Retrieval von Dokumenten - aufgrund der subjektiven Anforderungen des jeweiligen Benutzers - mit einem höheren Genauigkeitsgrad zu bestimmen, als das beim booleschen Retrieval der Fall ist.
- Die Retrievalsuche durch Anklicken einer graphischen Repräsentation zu aktivieren, was den Vorteil hat, daß der Benutzer die für ihn relevanten Dokumente identifizieren kann, auch wenn er nicht in der Lage ist, eine entsprechende Abfrage zu formulieren.

Sei LEX eine Menge von Lexemen, die vom Benutzer eingegeben wird. Für $A, B \in LEX$

und $\mathcal{O}(A)$ Menge der Texte, in denen eine Instantiierung von A vorkommt, kann man eine Relation, welche die intertextuelle gegenseitige kollokationelle Abhängigkeit zwischen A und B modelliert wie folgt definieren:

$$\text{conf}(A, B) = \frac{|\mathcal{O}(A) \cap \mathcal{O}(B)|}{|\mathcal{O}(B)|}$$

Falls in jedem Text, in dem eine Instantiierung von A vorkommt, auch eine solche von B vorkommt, dann wird $\text{conf}(A, B) = 1$ und falls A und B kein gemeinsames Vorkommen besitzen 0^1 . Dadurch wird eine

$$|LEX| \times |LEX| \text{ Matrix } CONF$$

berechnet, welche als Eintragungen die entsprechenden conf Werte hat. $\text{conf}(L_i, L_j)$ ist das $\langle i, j \rangle$ -te Element dieser Matrix und gibt die Gewichtung der Abhängigkeit $f_{ij} : L_i \rightarrow L_j$ an.

Wenn die Dokumentenbasis durch neue Dokumente augmentiert wird, kann die $CONF$ Matrix Neuberechnet werden, oder aber - falls die "Geschichte" der Dokumentenbasis mitberücksichtigt werden soll - kann conf zu einer inkrementellen Gewichtsbestimmung erweitert werden. Für die Neuberechnung von $\text{conf}(A, B)$ gibt es nach der Aufnahme eines weiteren Textes drei Fälle:

1. A und B sind im neuen Text erhalten.

In diesem Fall muß sowohl die Anzahl der gemeinsamen Vorkommen als auch diejenige der Vorkommen von B um 1 erhöht werden, d.h. $\text{conf}(A, B)$ wird

$$\text{conf}(A, B)_{\text{neu}} = \frac{|\mathcal{O}(A) \cap \mathcal{O}(B)| + 1}{|\mathcal{O}(B)| + 1}$$

Es gilt nun:

$$\text{conf}(A, B)_{\text{neu}} = \frac{\frac{|\mathcal{O}(A) \cap \mathcal{O}(B)| + 1}{|\mathcal{O}(B)| + 1}}{\frac{\text{conf}_{\text{alt}} + \frac{1}{|\mathcal{O}(B)|}}{1 + \frac{1}{|\mathcal{O}(B)|}}} =$$

2. Nur B kommt im neuen Text vor.

In diesem Fall wird nur $|\mathcal{O}(B)|$ um 1 erhöht. Analog zum ersten Fall haben wir:

$$\text{conf}(A, B)_{\text{neu}} = \frac{\text{conf}_{\text{alt}}}{1 + \frac{1}{|\mathcal{O}(B)|}}$$

3. Nur A kommt im neuen Text vor.

In diesem Fall gilt:

$$\text{conf}(A, B)_{\text{neu}} = \text{conf}(A, B)_{\text{alt}}$$

¹ conf ist analog zu Carnaps regulären Bestätigungsfunktion (Carnap, 1950) - hier in einem "Gebrauchsuniversum"- c zwischen Aussagen e und h .

Um die jeweils aktuelle $CONF$ Matrix zu bestimmen reicht es also wenn man die vorherige $CONF$ Matrix und die Anzahl der Vorkommen der einzelnen Lexeme gespeichert hat.

Aus der $CONF$ Matrix kann man nun die gewichteten Kanten eines gerichteten Graphen wie folgt bestimmen:

$$\text{conf}(A, B) \geq \text{conf}(B, A) \Rightarrow f : A \rightarrow B$$

$$\text{conf}(A, B) < \text{conf}(B, A) \Rightarrow f : B \rightarrow A$$

Dieser Graph kann als eine Hypertextstruktur aufgebaut werden, indem:

- jedes Lexem als eine *Karte*² implementiert wird und
- die Kanten, welche dieses Lexem mit anderen verbinden, als *buttons* implementiert werden, deren Name aus dem Ziellexem und der entsprechenden Gewichtung zusammengesetzt wird und die zu der Karte dieses Ziellexems führen.

Der Benutzer kann nun durch Anklicken der entsprechenden *buttons* durch das Netzwerk navigieren.

Um eine graphische Repräsentation des Netzwerks zu bekommen, wird ein *Hyperviewgenerator* eingesetzt, der aus der Hypertextstruktur eine *hyperview* erzeugt³. An diese *hyperview* werden die Dokumente angehängt, so daß man durch Anklicken eines Lexems (oder einer Menge von Lexemen) eine Hypertextstruktur bekommt, welche die Dokumente, in denen dieses Lexem vorkommt (oder diese Lexeme vorkommen), enthält.

4 Der Retrievalvorgang

Der Retrievalvorgang spielt sich in folgender Weise ab:

1. Eine Menge von gespeicherten Dokumenten wird in das System eingegeben, und es erfolgt sowohl eine Bestimmung der Topikhierarchien als auch der Kollokationsdependenzen. Die generierten Topikhierarchien werden an die *hyperview* des Kollokationsnetzes angehängt.

²Wir benutzen dabei die Terminologie von *hypercard*.

³Diese globale *hyperview* kann dann durch Interaktion mit dem Benutzer weiterhin restringiert werden bis der Benutzer eine Struktur erreicht, die für seine Zwecke ausreichend ist. Diese rekursive Restriktion basiert allerdings auf kategorientheoretische Formalismen, die hier nicht erörtert werden können, und die zu einer Lösung des Problems von "getting lost" führen ((Rieger & Thiopoulos, 1991) und (Bruza & Thiopoulos & van der Weide, 1992)). Diese Formalismen wurden im Rahmen einer Modellierung des phänomenologischen Bedeutungsbegriffs entwickelt ((Thiopoulos, 1990) und (Thiopoulos, 1992)).

2. Nachdem der Benutzer das Netz inspiziert hat, kann er für ihn interessante keywords anklicken. Damit gelangt er in einen korrespondierenden Topik, in dem die keywords vorkommen.
3. Er kann jetzt die entsprechende Topik-hierarchie inspizieren. Für den Fall, daß andere Teile dieser Topikhierarchie sein Interesse erregen, kann er durch Anklicken vorhandener *buttons* den Text durchmustern. Andernfalls kann er zum nächsten Text übergehen, genauer gesagt: zum nächsten Topik, in dem die spezifizierten keywords vorkommen. Von da aus hat er dieselben Möglichkeiten.
4. Während des Navigierens kann der Benutzer die für ihn interessanten Topiks in einer Hypertextstruktur ablegen, die das Ergebnis des Retrievalprozesses bildet.
10. Polanyi, Livia/Scha, Remko J .H. 1983. On the Recursive Structure of Discourse. In: Connectedness in Sentence, Discourse and Text. (Eds.) Ehlich, K. /van Riemsdijk, Henk, 141-178. Tilburg: University.
11. Rieger, B. *Unscharfe Semantik*. Frankfurt: Peter Lang, 1989.
12. Rieger, B. & Thiopoulos, C. A self organizing lexical system in hypertext. In *Proceedings of the first Conference on Quantitative Linguistics*. Trier, 1991. (demnächst)
13. Thiopoulos, C. Meaning metamorphosis in the semiotic topos. *Theoretical Linguistics*, 16, 2j3, 255-274, 1990.
14. Thiopoulos, C. Towards a logic of semiotic systems. In *Revue Mathematiques, Informatiques et Sciences Humaines*, December, 1991.

5 Bibliographie

1. Barrett, E. (Ed.). *Text, ConText, and HyperText*. MIT: 1988.
2. Bergler, Sabine 1991. The Semantics of Collocational Patterns for Reporting Verbs. Proceedings of the 5th *EACL Conference 1991* Berlin, 216-221
3. Bruza, P. D. & van der Weide T. P. Assessing the quality of hypertext views. *ACM SIGIR FORUM*, 24(3), 6-25, 1990.
4. Bruza, P. D. & Thiopoulos, C. & van der Weide, T. P. A categorial search in multi level hypermedia, 1992. (demnächst)
5. Carnap, R. *Logical foundations of probability*. Chicago: The University of Chicago Press, 1950.
6. Conklin, J. Hypertext: an introduction and survey. *Computer*, Vo1.20, No.9, September, 1987.
7. Furnas, G.W. Generalized fisheye views. *Human Factors in Computing Systems, CHI'86 Conference Proceedings*. Boston: 1986.
8. Jones, W. P. How do we distinguish the hyper from the hype in non-linear text? In: H.J. Bullinger & B. Shackel (Eds.). *INTERACT'87*. North Holland, 1987.
9. Nakhimovsky, Alexander/Rapaport, W.J.1988. Discontinuities in Narratives. Proceedings of COLING 1988 Budapest, 465-470
15. Weber, Heinz J. Faktoren einer textbezogenen Maschinellen Übersetzung: Satzstrukturen, Kohärenz- und Koreferenz-Relationen, Textorganisation. In: Batori, J. S. & Weber, H. J.(Eds.). *Neue Aufsätze in Maschineller Übersetzung: Wissensrepräsentation und Textbezug*. Tübingen: Niemeyer, 1986.
16. Weber, Heinz J. Text-Serien, Text-Ebenen und Text-Instanzen in journalistischen und amtlichen Nachrichten In: Fachsprachlichkeit in Texten. (Ed.) Kalverkämper, H., Tübingen: Narr. (demnächst)