

Kohärenzrelationen in Hypertexten¹

Rainer Hammwöhner

Universität Konstanz

Fachgruppe Informationswissenschaft

Inhalt

1. Einleitung
2. Textlinguistische Grundlagen
3. Automatischeinhaltserschließung durch partielles Parsing: Das TOPIC-System
4. Hypertext-Links in TWRM-TOPOGRAPHIC
 - 4.1. Bestätigung
 - 4.2. Elaboration
 - 4.3. Inkompatibilität
 - 4.4. Rollenwechsel
 - 4.5. Ähnlichkeit
 - 4.6. Kontrast
5. Ausblick
6. Literatur

Referat

In diesem Beitrag werden getypte Hypertext-Links als Analoga zu den Kohärenzrelationen linearer Texte eingeführt. Exemplarisch werden einige derartiger Relationen vorgestellt, wie sie zur Verbesserung der Leistung des interaktiven graphischen Retrieval-Systems TWRM-TOPOGRAPHIC spezifiziert wurden.

Abstract

This papers deals with typed hypertext links which are introduced as analogues to the relations of coherence which may be found in linear texts. Some relations are presented as examples which were specified to improve the performance of the interactive graphical retrieval-system TWRM-TOPOGRAPHIC.

1 Einleitung

Eine Grundlage für einen flexiblen, die übliche lineare Textform sprengenden Umgang mit Textinformation bilden die in den letzten Jahren entwickelten Hypertext-Systeme, die eine

¹ Dieser Text ist erschienen in: Josef Herget, , Rainer Kuhlen (Hrsg) Pragmatische Aspekte beim Entwurf und Betrieb von Informationssystemen. Proc. des 1. Int. Symp. für Informationswissenschaft. UVK, 1990, S. 44-57.



Dieser Text ist unter der folgenden Creative Commons Lizenz lizenziert: Attribution-NonCommercial-NoDerivs 2.0 Germany (<http://creativecommons.org/licenses/by-nc-nd/2.0/de/>).

netzwerkartige Strukturierung von Textmengen mit graphisch interaktiven Präsentationsmöglichkeiten verbinden. Die Flexibilität dieses neuen Mediums läßt es, unter anderem, auch als geeigneten Informationsträger zur Distribution von Fachinformation erscheinen. Gegenüber konventionellen Retrievalsystemen tritt insbesondere der Vorteil hervor, daß Texte, bzw. Textsegmente durch inhaltliche Vernetzung in den Kontext einer größeren Textmenge eingebettet werden können. Erforderlich ist aber eine kohärente Strategie zum sukzessiven Hypertextaufbau (aus Fachtexten), die Verfahren zur Auswahl der Hypertext-Einheiten, eventuell durch Segmentierung der Texte, und zur inhaltlichen Verknüpfung dieser Textfragmente einschließt. Das bezieht sich einerseits auf den semantischen Gehalt intra- und intertextueller Beziehungen, die durch Hypertext-Links ausgedrückt werden sollen. Wie im nächsten Abschnitt ausgeführt wird, kann hier auf Modelle der Textlinguistik zurückgegriffen werden. Andererseits stellt sich auch das Problem der intellektuellen Kontrollierbarkeit eines aus großen Textmengen sukzessive aufzubauenden Hypertexts. Nach unserem Dafürhalten kann Konsistenz in der Vergabe textübergreifender Hypertext-Links nur durch eine Automatisierung dieses Vorgangs erreicht werden. Dementsprechend beruhen die im Hauptabschnitt dieser Arbeit vorgestellten intertextuellen Relationen auf Textrepräsentationen, wie sie mit einem Textparser (der TOPIC-Parser), der dem gegenwärtigen Stand der Kunst entspricht, aus Fachtexten erschlossen werden können.

2 Textlinguistische Grundlagen

Hypertexte sind nie auf vollständige Rezeption hin ausgelegt, so daß die Auswahl und endgültige Anordnung der Inhalte in den Hypertext-Pfaden, den vom Leser rezipierten Konkretisierungen des Hypertextes, erst von diesem endgültig bestimmt werden. Diese gegenüber konventionellen Texten für den Leser gewonnenen zusätzlichen Freiheitsgrade lassen sich im Rahmen zweier alternativer Modelle interpretieren:

- Ein Hypertext ist ein prästabilisierter Text, aus dem durch eine Stoffreduktion auf das jeweils Relevante und durch Linearisierung **ein Text** entsteht. In diesem Fall sind die Modelle der Textlinguistik auf das entstehende Gebilde direkt übertragbar.
- Ein Hypertext ist eine **Menge von Texten**, die durch inhaltliche Vernetzung in ihrer **Intertextualität** repräsentiert werden. Für derartige textübergreifende Strukturen gibt es bisher allerdings keine operationalisierbaren Formalisierungen.

In diesem Spannungsfeld zwischen Textualität und Intertextualität befindet sich ein aus deskriptiv expositorischen Fachtexten — im weiteren Journalartikel aus der Mikrocomputertechnologie — zum Zweck des Information Retrieval aufgebauter Hypertext zwischen den beiden Polen. Es ist zwar richtig, daß der Aufbau des Hypertexts aus disparaten Texten zunächst für ein Überwiegen intertextueller Beziehungen spricht. Grundlage zur Lösung eines Informationsproblems sind nicht einzelne Texte mit ihrer jeweils unzureichenden Information, sondern der von diesen Texten gebildete Inter- bzw. Hypertext. Der Umgang von Lesern mit Fachjournalen (McKnight et al. 89), die derartige Texte enthalten und daher die konventionelle Repräsentation des Intertexts bieten, legt allerdings die Vermutung nahe, daß der zwar an der vorgegebenen Dokumentenstruktur orientierte, aber auswählende, nicht dem Textverlauf folgende, Textgrenzen überspringende Lesestil Intertextualität² in Textualität umdeutet, so daß Fragmente aus unterschiedlichen Texten unter dem Gesichtspunkt einer thematischen Vollständigkeit und inhaltlichen Kohärenz neu organisiert werden. Diese Betrachtungsweise ermöglicht es, die entstehenden

² Selbstverständlich ist die literarische Auffassung von Intertextualität, die formal kaum zu erfassende Phänomene wie Parodie umfaßt, erheblich umfassender als die im Kontext dieser Arbeit vertretene.

Strukturierungsprobleme im Lichte textlinguistischer Modelle zu betrachten und bedingt damit zwangsläufig die Frage nach den strukturellen Vorbedingungen von Textualität, die auf drei Ebenen anzusiedeln sind (Hatakeyama et al 85):

- *Konnektivität* entsteht, wenn sich über eine Folge von Textsegmenten mit Hilfe von zunächst arbiträren Texteigenschaften ein roter Faden konstruieren läßt. Diese Eigenschaften können sowohl auf der Ebene der Phone (Versmaß, Reim etc.), wie auch der Syntax (Konnektiva), Semantik (sinn-semantische Relationen) und Pragmatik angesiedelt sein. Für uns werden im weiteren vor allem die referentiellen Relationen von Bedeutung sein, die in referenzidentische Relationen, die Begriffswiederholungen, Pro-Formen und lexikalisch-referenzielle Relationen wie Hyponymie und Hypernymie umfassen, und referenzkonforme Thesaurusrelationen, die durch Erwähnung von Aspekten schon referenzierter Begriffe bzw. Objekte entstehen, aufgeteilt werden.
- *Kohäsion* ist eine striktere Form der Konnektivität, die sich durch das zusätzliche Bestehen sinn-semantischer Relationen und adäquater thematischer Progressionen konstituiert.
- *Kohärenz* ist trotz zahlreicher textlinguistischer Untersuchungen ein relativ opaker Begriff (Horányi 85). Gemeinsam ist die Auffassung, daß Kohärenz nicht allein textimmanent zu begründen ist, sondern — auch in Analogie zu anderen Strukturen, wie z.B. Bildern (Dorfmueller-Karpusa/Dorfmueller 85) — von Eigenschaften der realen/fiktionalen Welt, Erwartungen des Lesers/Betrachters usw. abhängt (van de Velde 85), also pragmatische Aspekte hat. Eine Integration von pragmatischen und referentiell semantischen Aspekten der Kohärenz gibt Heydrich 89, der Kohärenz durch Beziehungen zwischen relevanten Objekten bzw. Situationen konstituiert sieht.

Kohäsion und Kohärenz betreffen sowohl oberflächen-syntaktische wie auch semantische Strukturen. Eine Übernahme der für lineare Texte eingeführten oberflächen-syntaktischen Kohäsions- oder Kohärenzindikatoren in Hypertexte ist nicht möglich. Dieses Defizit muß wegen der großen Bedeutung von Strukturhinweisen für das Textverstehen (Kieras 82) durch Einführung neuer Stilmittel in die Hypertext-Präsentation (Charney 87) ausgeglichen werden. Insbesondere ist eine Ausweitung der kohäsions- und kohärenzstiftenden Funktion graphischer Elemente über das in linearen Texten übliche Maß (Liebsch/Werchosch 88) hinaus anzustreben. Die text-semantische Ebene wird schon in linearen Texten von nicht-linearen Kohärenzrelationen gebildet, so daß von einer zweidimensionalen Struktur von Texten (Gulich/Raible 77 pp. 51-55) gesprochen werden kann. Diese Relationen finden ihre Entsprechung in den Links der Hypertexte, so daß der Übergang zum Hypertext eine Explikation von Strukturen mit sich bringt, die schon in linearen Texten implizit enthalten sind. In diesem Zusammenhang ist insbesondere auf die Arbeiten von Hobbs (z.B. Hobbs 85) und Mann/Thompson (z.B. Mann/Thompson 88) zu verweisen, in denen ein Kanon von Kohärenzrelationen vorgestellt wird.

Für die automatische Konvertierung von Textmengen in Hypertexte stellen sich dementsprechend insbesondere folgende Fragen.

- Wie können Texte fragmentiert werden; so daß die entstehenden Texteinheiten in sich konnex bleiben (s. Hammwöhner 90)? Ergebnisse der Textlinguistik und der Psychologie legen eine Segmentierung auf dem Niveau von Paragraphen nahe.
- Welche Verbindungen zwischen den Hypertext-Einheiten — auch solchen, die ursprünglich zu unterschiedlichen Texten gehörten — lassen sich etablieren, so daß eine inhaltsorientierte Navigation ermöglicht wird (s.u.)?

- Welche globale Strukturen können konstituiert werden, die dem Leser bei der Hypertext-Navigation als Orientierungshilfe dienen und Kohäsion und Kohärenz der gelesenen Folge von Hypertext-Einheiten garantieren (s. Hammwöhner 90)?

3 Automatische Inhaltserschließung durch partielles Parsing: Das TOPIC-System

Für Inhaltserschließung und Repräsentation im Rahmen der automatischen Konstruktion von Hypertexten lassen sich folgende Anforderungen benennen:

- Sowohl die Segmentierung von Texten als auch die Vernetzung der entstehenden Textfragmente basieren in erster Linie auf semantischen Kriterien, weshalb die syntaktische Struktur der Texte weitgehend vernachlässigt werden kann.
- Zur Bestimmung der referentiellen Struktur des Textes, entscheidend für den Aufbau der Makrostruktur, müssen textlinguistische Phänomene, wie Anaphora, pronominale Referenz etc. behandelt werden können.
- Kriterien zur Ableitung der Salienz der im Text auftretenden Konzepte beruhen auf Kohärenzphänomenen, z.B. der thematischen Progression (Janoš 79).
- Die Unbestimmtheit des sich sukzessiv aufbauenden hypertextuellen Kontextes erlaubt keine fokussierte auf ein bestimmtes Interesse ausgerichtete Inhaltserschließung.
- Die Vernetzung der Texteinheiten soll im Sinne einer kohärenten thematischen Entwicklung während der Hypertext-Navigation, auf essentiellen, nicht akzidentiellen Inhalten beruhen, insofern ist eine Kondensierung der Inhalte anzustreben.
- Die Notwendigkeit, größere Textmengen bearbeiten zu können, erfordert eine effiziente Inhaltserschließung.

Diese Rahmenbedingungen stellen die Analyse-Komponente des zu entwerfenden Hypertext-Systems in den Kontext des wissensbasierten Abstracting, wie es von dem in Konstanz entwickelten TOPIC-System geleistet wird. Die Übernahme der methodischen Grundlagen von TOPIC³, insbesondere des Repräsentationsformalismus, ermöglicht es, den Aspekt der Textanalyse nur cursorisch zu streifen und, die Analyseergebnisse voraussetzend, auf die Definition von Kohärenzrelationen einzugehen.

Das TOPIC-System ist ein Textanalyse und -kondensierungssystem, das Zeitschriftenartikel eines eingeschränkten Diskursbereichs im Sinne eines indikativ-informativen Referats inhaltlich erschließt, indem Thematik und wichtige Fakten eines Textes bestimmt werden. Im Gegensatz zum traditionellen Abstracting und auch zu derzeit verfügbaren automatischen Abstracting-Systemen (z.B. DeJong 82, Fum et al 82, Tait 85) – eine Übersicht gibt Hutchins 87 – besteht die Leistung von TOPIC nicht: in der Generierung eines auf ein vorgegebenes Interessenprofil ausgerichteten natürlichsprachlichen Abstracts. Vielmehr wird der Text in thematisch kohärente Teiltexthe zerlegt, deren thematische Struktur jeweils durch Themenbeschreibungsglyphen (Reimer/Hahn 88) repräsentiert wird.

Den speziellen Anforderungen an die Inhaltserschließung zum Zweck des Abstracting bzw. Indexing — Effizienz und Robustheit bei Textverstehen begrenzter Tiefe — wird TOPIC durch Einsatz eines konzept-orientierten partiellen semantischen Parsers gerecht. Die linguistische Kompetenz des Parsers ist objektorientiert in Form von Wort-Experten (Hahn

³ Eine ausführliche Beschreibung von TOPIC geben Hahn/Reimer (86a,88).

87) spezifiziert. Im Unterschied zu anderen Ansätzen (s. Reddig 84) sind die Wortexperten in TOPIC nicht strikt wortorientiert, sondern beschreiben, basiert auf dem Konzeptwissen⁴ einer frame-orientierten Wissensbasis, in erster Linie satzübergreifende textlinguistische Phänomene, insbesondere referenzieller Art, wie z.B. Anaphern und Proformen. Der Frame-Ansatz unterstützt aufgrund der objektzentrierten Modellierung (Minsky 75), die besonders zur Darstellung referenzieller Relationen geeignet ist, die einfache Formulierung von Wortexperten, die Referenzketten verfolgen und damit Anaphern und Proformen auflösen, aber auch thematische Progressionsmuster aufdecken können (Hahn 90). Dabei entsprechen referenz-konforme Thesaurus-Relationen den Slot- und Slot-Eintrags-Relationen des Frame-Modells, während referenz-identische Relationen einer durch die Slot-Struktur von Frames induzierten Spezialisierungshierarchie entsprechen. Diese merkmalsabhängige Definition der Spezialisierung (für eine formale Definition s. Reimer 89) ist ein Spezifikum des eigens für TOPIC entwickelten Repräsentationsmodells FRM⁵ (Reimer 89). Im Gegensatz zu anderen frame-orientierten Repräsentationssprachen, z.B. KL-ONE (Brachman/Schmolze 85), verfügt FRM weiterhin über eine weitgehende, durch modellinhärente Integritätsregeln gesteuerte Integritätskontrolle für Wissensbasen, die bei Änderungen der Wissensbasis durch Knowledge-Engineering, aber auch durch Konzeptlernen während der Analyse, eine fortwährende Validität der Wissensbasis garantiert.

Als Ergebnis der Textanalyse liegt eine modifizierte Wissensbasis vor, die die thematische Struktur des untersuchten Textabschnitts wie folgt repräsentiert:

- Die Salienz der auftretenden Begriffe wird durch die Zuweisung eines Aktivierungsgewichts, das der nach Auflösung von Anaphern und Proformen ermittelten Erwähnungshäufigkeit proportional ist, wiedergegeben. Derartige Gewichtungen können bei Frames, Slots und Einträgen vorgenommen werden.
- Werden im Text einem Konzept Merkmalsausprägungen zugewiesen, so wird eine entsprechender Slot eintrag in der Wissensbasis vorgenommen.
- In besonderen, allerdings sehr eingeschränkten Fällen, kann ein nicht in der Wissensbasis repräsentiertes Konzept aus dem Text erworben werden. Die Verbesserung dieser Lernfähigkeit ist Gegenstand aktueller Forschungsarbeit (Reimer 90).

An die Analyse des Textes schließt sich eine Kondensierungsphase an, die unterdurchschnittliche Aktivierungen, die vermutlich akzidentiell sind, eliminiert. Zusätzlich zum Aktivierungswert wird bei der Kondensierung auch die/ Existenz von Slot-Einträgen als ein Indikator für die Bedeutung eines Konzepts berücksichtigt. Die somit aus der Textanalyse resultierenden Textwissensbasen bilden die Grundlage für die Berechnung von Kohärenzrelationen zwischen den entsprechenden Textabschnitten.

Das folgenden Beispiel vermittelt einen Eindruck von der Ausnutzung semantischer Strukturen für Merkmalszuweisung, Auflösen von Anaphern und einfaches Konzeptlernen. In einem Artikel sei die folgende Formulierung enthalten:

Der IBM-PC ist ein häufig verkaufter Personalcomputer.....Dieser Rechner verfügt über einen 8080-Mikroprozessor, der

Die salienten Begriffe *Personal-Computer*, *Rechner* und *8088-Mikroprozessor* seien, wie der Abbildung 1 zu entnehmen ist in eine Frame-Struktur eingeordnet, während der Begriff *IBM-*

⁴ Die Textsorte erlaubt eine Begrenzung der Analyse auf Nominalphrasen. Deshalb sind zur Zeit nur Nomina im Lexikon enthalten.

⁵ Frame-Repräsentations-Modell

PC nicht in der Wissensbasis enthalten ist. Durch das Vorwissen, daß *PC* einen Personal-Computer bezeichnet, *IBM* ein Hersteller ist, kann ein Frame für das Kompositum *IBM-PC* bereitgestellt werden, der von dem Prototyp *Personal-Computer* abgeleitet ist und über einen Eintrag in den Slot *Hersteller* verfügt. Indem die Referenz auf den Begriff *Rechner* durch referenzidentische Relationen auf *IBM-PC* (Synonymie, Unterbegriff) zurückgeführt werden kann, wird die Zuweisung eines weiteren Slot-Eintrags ermöglicht, dementsprechend wird eine Erhöhung des Aktivierungsgewichts nicht beim Konzept *Rechner* sondern bei *IBM-PC* vorgenommen.

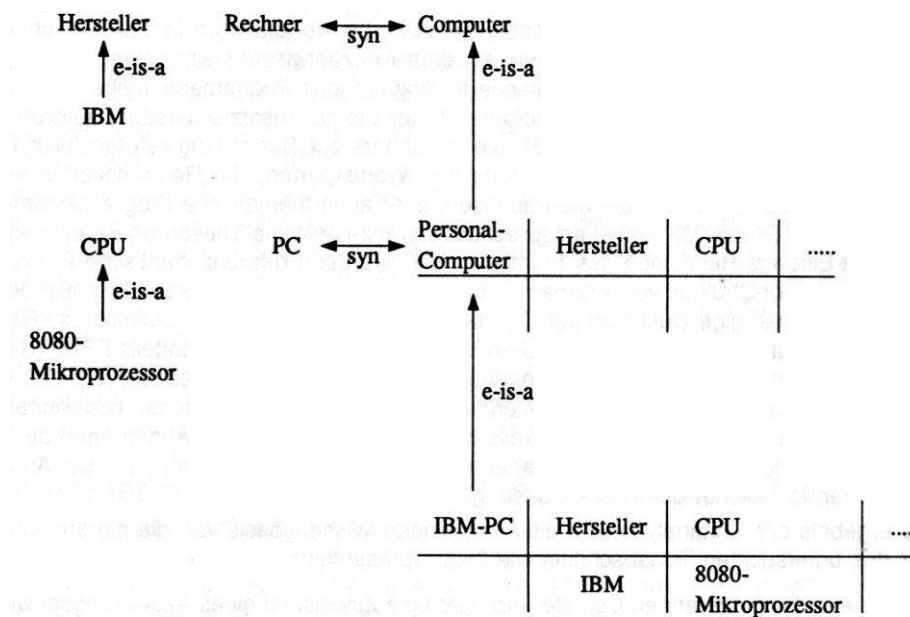


Abbildung 1 Textwissensfragment

4 Hypertext-Links in TWRM-TOPOGRAPHIC

In diesem Abschnitt werden zweistellige Relationen auf der Menge der Textrepräsentationen eingeführt, die als Hypertext-Links die Kohärenz des Hypertext-Korpus konstituieren. Die folgenden Definitionen geben die semantischen Vorbedingungen für das Bestehen der jeweiligen Relationen⁶ wieder, deren pragmatische Aspekte erfaßt werden, indem sie in Verwendungszusammenhänge, d.h. Textschemata eingeordnet werden⁷. Die Relationen sind inhaltlich an den von Hobbs und Mann/Thompson formulierten Kohärenzrelationen orientiert, aber in erster Linie auf den Diskursbereich zugeschnitten, so wurde z.B. auf temporale und kausale Verknüpfungen verzichtet, da diese in den weniger argumentativen als deskriptiven Texten keine bedeutende Rolle spielen. Weitere Einschränkungen resultieren aus der Begrenztheit der zur Verfügung stehenden Inhaltserschließungs- und Repräsentationsverfahren. Inwieweit z.B. eine Textpassage den Leser zu motivieren vermag – Motivation gehört zu den rhetorischen Relationen der RST (Mann/Thompson 88) – kann mit den hier eingesetzten Mitteln nicht erschlossen werden, führt aber beim konstruktiven Gebrauch der Begrifflichkeiten auch eher in die Richtung konversationaler Diskursmodellierung, wie sie von Thiel (90) für das Information Retrieval vorgeschlagen wird. Andererseits sind einige unmittelbar auf den Diskursbereich zugeschnittene Relationen eingeführt, wie die Gegenüberstellung vergleichbarer Objekte oder die Berücksichtigung

⁶ Der Link-Taxonomie von DeRose (89) folgend sind sie deshalb als intensionale Links zu bezeichnen.

⁷ Eine formale Spezifikation der Kohärenzrelationen, sowie der Textschemata gibt Hammwöhner 90.

unterschiedlicher Rollen. Der Kanon der hier definierten Relationen soll und kann nicht vollständig sein, sollte aber zur Darstellung der Funktion von Hypertext-Links in dem vorgeschlagenen Hypertext-Modell ausreichen.

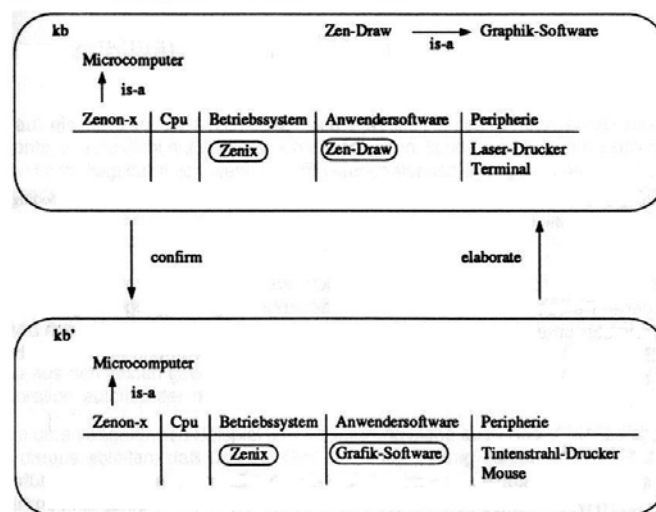
Die im folgenden definierten Relationen beruhen auf Eigenschaften der Textrepräsentationen, insbesondere auf der Einordnung von Objekten, die im folgenden als Frames modelliert werden, in die Konzepthierarchie, auf den zu den Objekten akquirierten Fakten und auf der thematischen Relevanz (van Dijk 79) der angesprochenen Begriffe.

4.1 Bestätigung

Durch Einführung kontrollierter Redundanz z.B. durch Textsegmente, die bisher getroffene Feststellungen bestätigen, kann die Akzeptanz eines (Hyper-)Textes durch den Leser gesteigert werden. Eine solche Bestätigung kann erfolgen durch:

- Wiederholung von Faktenangaben,
- Subsumption von Fakten unter einen umfassenderen Begriff oder durch
- Aufsuchen eines Faktums von dem auf das zu bestätigende Datum geschlossen werden kann.

Zusätzlich dürfen in den betroffenen Texteinheiten keine inkompatiblen Angaben (*conflict-Relation*, s. Abschnitt 4.3) enthalten sein, selbst wenn sie das abzusichernde Faktum nicht unmittelbar betreffen.



Die Abhängigkeit zwischen Merkmalsausprägungen ergibt sich dabei aus den, den Slots zugeordneten Integritätsregeln. Das in Abb. 2 dargestellte Beispiel zeigt, daß eine bestimmte Anwendersoftware nur dann einem Rechner zugesprochen werden kann, wenn das Betriebssystem, für das sie entworfen wurde, mit dem des Rechners übereinstimmt.

Eine andere Form der Bestätigung ergibt sich aus einer Wiederholung des Sachverhalts unter veränderter Perspektive durch Umkehrung der zugehörigen Aggregierungsrelation. Wurde zuvor ein Computer unter dem Gesichtspunkt seiner Bauteile untersucht, wird nun das Vorkommen eines Bauteils in verschiedenen Rechnern, darunter auch der vorgegebene, betrachtet. Damit ergibt sich zusätzlich ein Wechsel des Diskurs-Fokus.

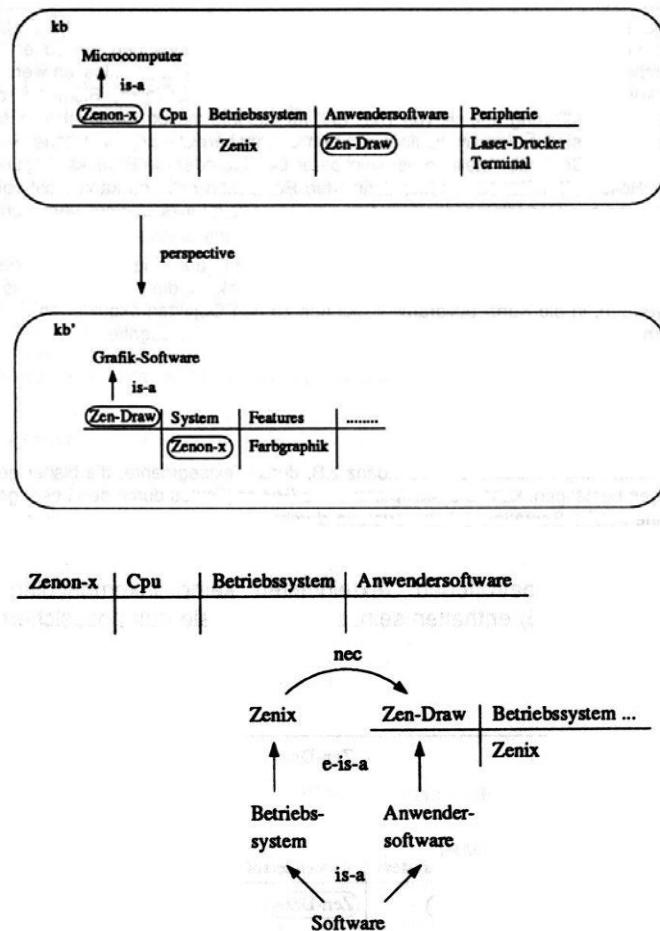
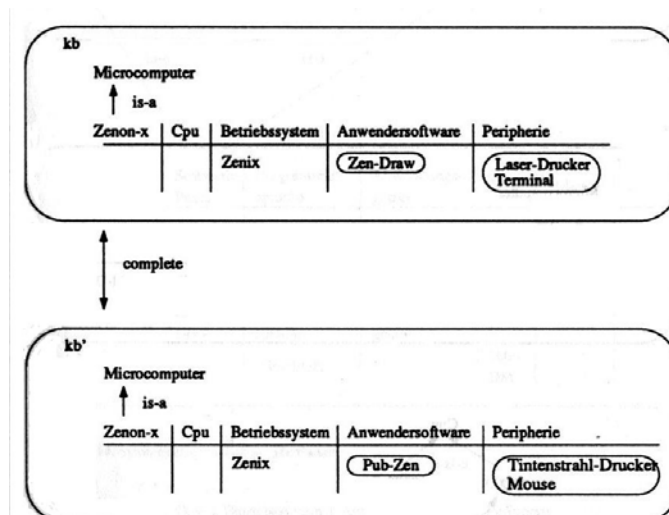


Abbildung 2 Dem Slot *Anwendersoftware* des Frame *Microcomputer* ist eine Integritätsbedingung zugeordnet, welche die Gleichheit der Ausprägungen des Merkmals *Betriebssystem* fordert.

4.2 Elaboration

Im Gegensatz zur Bestätigung, die ein retardierendes Element in den Text einführt, wird die Entwicklung des (Hyper-)Textes durch Elaboration (*elaborate*) im Sinne eines lokalen Diskursziels vorangetrieben, indem allgemeine Begriffe durch konkretere ersetzt werden oder die Konsequenzen vorgegebener Prämissen aufgezeigt werden. Von dieser Form der Elaboration, die auf eingeführte Fakten Bezug nimmt, läßt sich die Vervollständigung (*complete*) unterscheiden, die völlig neue Angaben zu einem Objekt erfordert.



4.3 Inkompatibilität

Eine auf ein Konzept zu beziehende Inkompatibilität zwischen den Ausführungen zweier Textfragmente, wie sie nicht nur durch unrichtige Angaben, sondern auch durch Zeitversionen etc. vorkommen kann, liegt dann vor, wenn in ihnen Merkmalsausprägungen zugewiesen werden, die aufgrund modell- bzw. domänenspezifischer Integritätsregeln nicht gemeinsam in einem Frame auftreten dürfen. Zur Überprüfung der Kompatibilität von Einträgen in nicht-terminalen Slots kann hier nicht allein der Frame-Name herangezogen werden, sondern es muß die Frame-Struktur überprüft werden.

Das Bestehen dieser Relation zwischen zwei Texteinheiten kann, muß aber nicht zwangsläufig auf Fehlinformation in einer der beiden hinweisen. Es kann sich auch um Beschreibungen zweier nicht kompatibler Ausprägungen einer Objektklasse handeln. In beiden Fällen können die aus den Texten gewonnenen Fakten nicht im Sinne einer gegenseitigen Bestätigung oder Elaboration aufeinander bezogen werden.

In dem unten angeführten Beispiel (zur Erläuterung siehe auch Abb. 3) läßt sich ein Konflikt einerseits daraus ableiten, daß *Cpu* als Slot nur einen Eintrag zuläßt, andererseits aber auch daraus, daß eine Integritätsregel die Übereinstimmung der Angaben zur *Cpu* von *Betriebssystem* und *Rechner* verlangt.

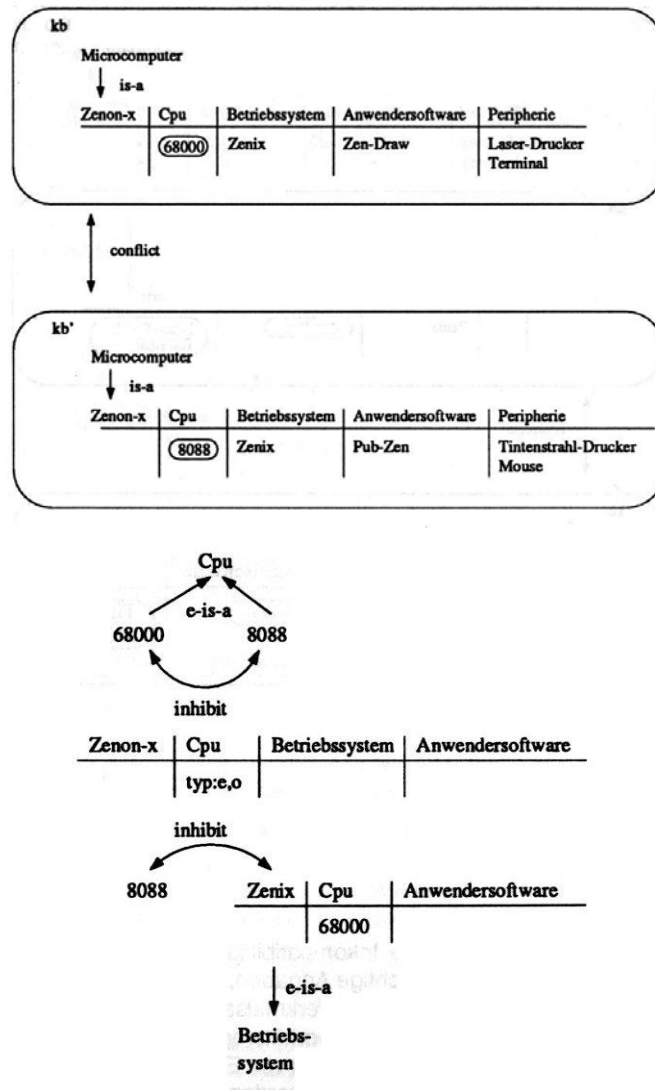


Abbildung 3 Dem Slot *Betriebssystem* des Frame *Zenon-x* ist eine Integritätsbedingung zugeordnet, welche die Gleichheit der Ausprägungen des Merkmals *Cpu* fordert.

4.4 Rollenwechsel

Diese Relation bezeichnet die Diskussion eines Objekts unter unterschiedlichen Perspektiven. Diese Sichtweisen lassen sich unterschiedlichen Oberbegriffen in der Spezialisierungshierarchie zuordnen. Ein Perspektivenwechsel zwischen zwei Texteinheiten findet also genau dann statt, wenn in jeder von ihnen Aspekte des Objekts diskutiert werden, die sich nur einem dieser Oberbegriffe zuordnen lassen. Behandelt z.B. (s. Abb. 4) eine Textpassage ein Software-Paket unter dem Gesichtspunkt des Programmierstils, während eine andere auf Preis und Marktchancen abhebt, so liegt ein Wechsel der Perspektive vor, der durch den Übergang vom Oberbegriff *Software* zum Oberbegriff *Produkt* gekennzeichnet ist.

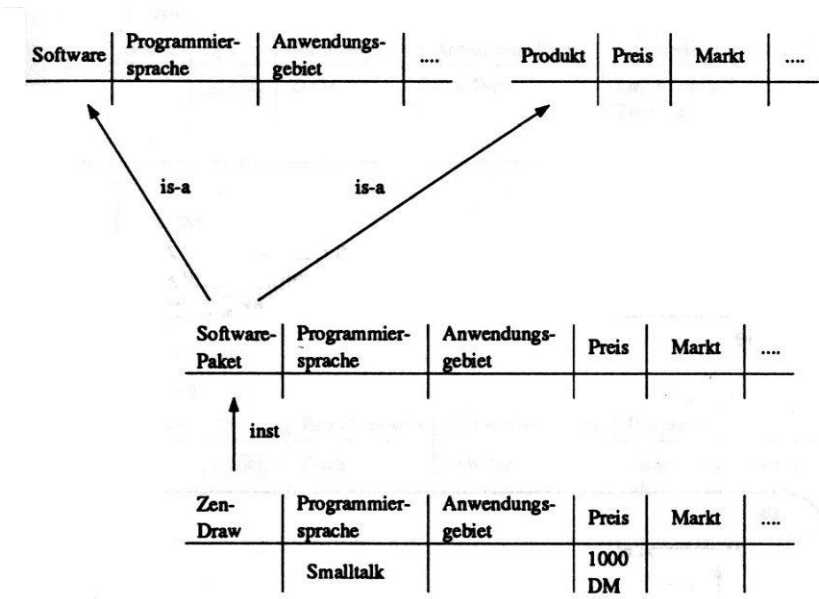
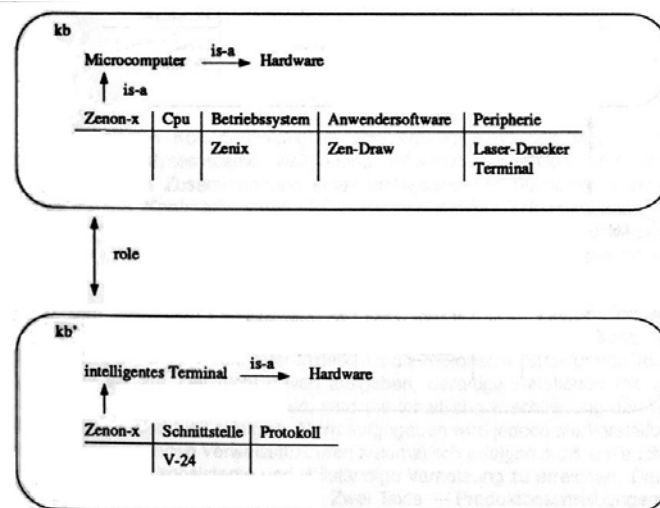


Abbildung 4 Ein Wissensfragment, welches das Konzept *Software-Paket* einerseits als *Software* und andererseits als *Produkt* und damit in zwei Rollen modelliert.

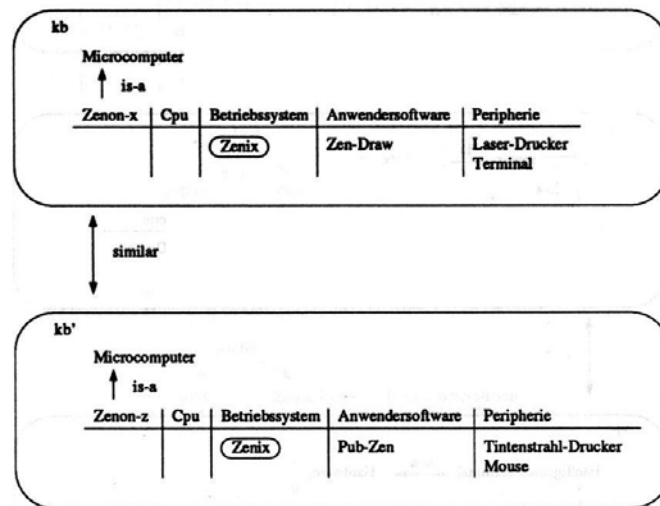


Im allgemeinen werden solche Perspektiven in einem gemeinsamen Unterbegriff integriert (*Software-Paket*). Liegt aber ein ungewöhnlicher Verwendungszusammenhang vor (Mikrorechner als Arbeitsplatzcomputer, aber auch als intelligente Terminals), so ist möglicherweise keine integrierende Sicht modelliert. In diesem Fall kann ein Konzept in zwei Text-Repräsentationen verschiedenen Prototypen zugeordnet werden. Diese Behandlung von "Rollen" als nur lokal gültige Einordnung in eine Spezialisierungshierarchie ist vielleicht formal weniger stringent als andere Ansätze zu diesem Problem, die z.B. eigene Modellkonstrukte für Rollen vorsehen (Reimer 86), aber für den Einsatz im automatischen

Aufbau von Hypertexten ist vor allem die Flexibilität der vorliegenden Lösung entscheidend, die es erlaubt, unterschiedliche Rollen evtl. auch ohne eine letztlich starre Vormodellierung zu erfassen.

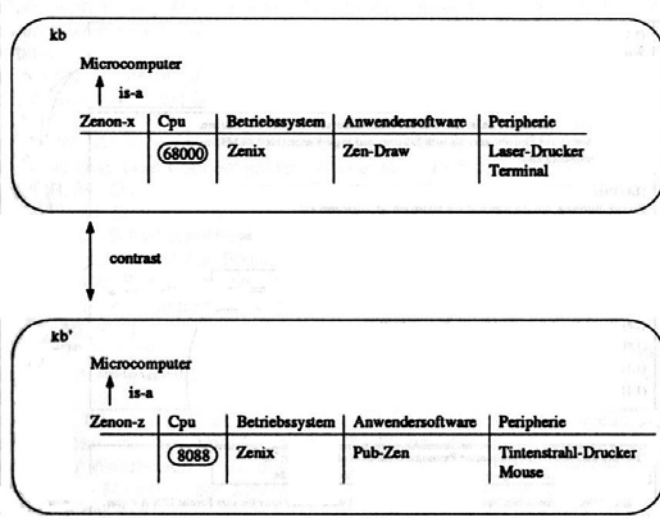
4.5 Ähnlichkeit

Übereinstimmung zweier Instanzen eines Prototyps bezüglich der zu einem Merkmal akquirierten Fakten ist eine sinnvolle Voraussetzung für eine vergleichende Gegenüberstellung insbesondere im Zusammenhang mit einer Kontrastierung (s. Abschnitt 4.6) hinsichtlich zweier anderer Merkmale.



4.6 Kontrast

Sind zwei Instanzen eines Prototyps Fakten zugeordnet, die, analog zur Konflikt-Relation (s. Abschnitt 4.3), aufgrund modellinhärenter oder domänenspezifischer Integritätsregeln inkompatibel sind, so sind die Voraussetzungen für eine kontrastive Gegenüberstellung von Texteinheiten gegeben, die insbesondere bei Übereinstimmung hinsichtlich anderer Merkmale informativ sein kann (s. Abschnitt 4.5).



5 Ausblick

Die oben definierten Kohärenzrelationen sind entsprechend den Möglichkeiten derzeit verfügbarer Volltext-Analysesysteme weitgehend referenziell orientiert und ohne tieferen

rhetorischen Gehalt. Im Zusammenhang eines umfassenderen Textschemas (Hammwöhner 90), das Elaborationen, Kontrastierungen etc. einem bestimmten Diskursziel zuordnet, das hier in Beschreibung oder Vergleich von Produkten aus dem Bereich der Mikrocomputertechnologie bestehen kann, ermöglichen sie dennoch Informationsdienstleistungen, wie sie von konventionellen Retrievalsystemen nicht angeboten werden können.

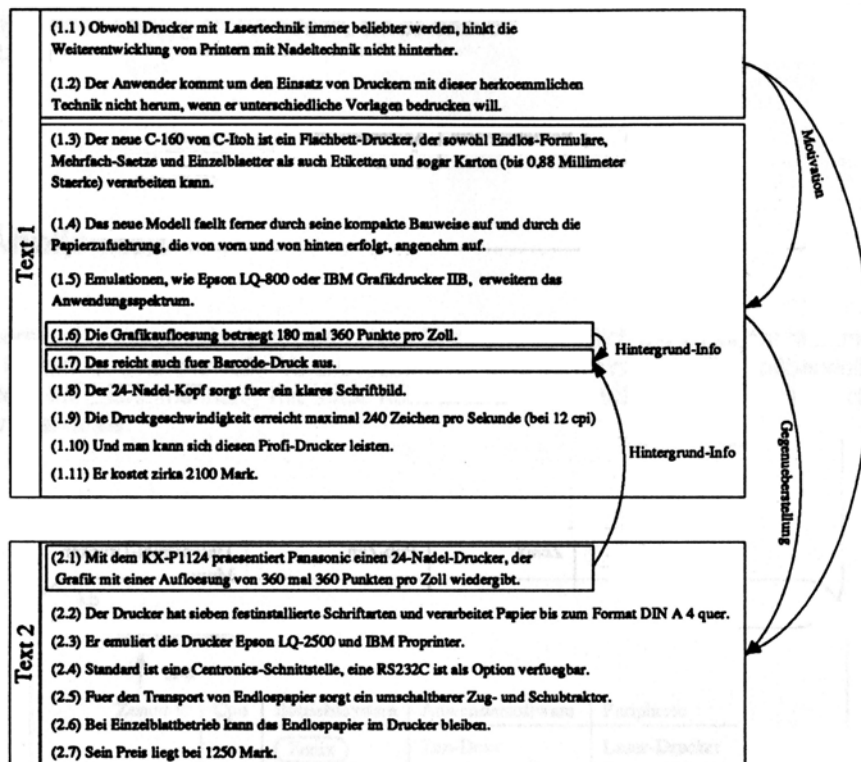


Abbildung 5 Zwei kurze Fachtexte, die Produktbeschreibungen von Nadeldruckern enthalten. Die jeweils textinternen Relationen werden intellektuell erschlossen, während die textübergreifenden Relationen automatisch vergeben werden.

Die über den aktuellen Zeitpunkt hinaus am Lehrstuhl für Informationswissenschaft geplante Forschung auf dem Gebiet Hypertext schließt aber die Erfassung komplexerer Kohärenzrelationen, die weniger die referenzielle als die rhetorische Struktur von Texten widerspiegeln, ein. Da wir zur Zeit nicht davon ausgehen, derartige Relationen mit vertretbarem Aufwand automatisch erkennen zu können, wird die inhaltliche Erschließung der Texte zumindest vorübergehend intellektuell erfolgen. Nicht aufgegeben wird jedoch die Vorstellung, daß der Aufbau der textübergreifenden Verweisstrukturen automatisch erfolgen muß, um auch in umfangreichen Hypertexten eine konsistente und vollständige Vernetzung zu erreichen. Das Beispiel in Abbildung 5 illustriert diese Vorgehensweise: Zwei Texte — Produktbeschreibungen von Nadeldruckern — werden intellektuell inhaltlich erschlossen, das heißt das eine Segmentierung in funktional aufeinander bezogene Texteinheiten stattfindet, deren propositionaler Gehalt, Thematik und Funktion für den Gesamttext angegeben werden. Der erste Abschnitt von Text 1 z.B. dient als Motivation sich trotz des Booms der Laser-Drucker-Technik mit Nadel-Druckern zu befassen, d.h. den weiteren Text zu lesen. Unter Vorgabe von Verallgemeinerungsbedingungen für die rhetorische Funktion *Motivation*, wie z.B. thematische Übereinstimmung des Zielabschnitts, läßt sich eine intertextuelle Beziehung automatisch ableiten, d.h. der erste Abschnitt von Text 1 kann analog auf den Text 2 bezogen werden.

6 Literatur

1. **Brachman, R.J. / Schmolze, J.G. 85:** An Overview of the KL-ONE Knowledge Representation System. In: Cognitive Science, Vol. 9, 1985, pp. 171-216.
2. **Charney, D. 87:** Comprehending Non-linear Text: The Role of Discourse Cues and Reading Strategies. In: Proc. of the Hypertext '87 Conference, 13.-15.11.87, Univ. of NC, Chapel Hill, NC, 1987, pp. 109-120.
3. **DeJong, G. 82:** An Overview of the FRUMP System. In: Lehnert, W. / Ringle, M. (eds): Strategies for Natural Language Processing. Hillsdale/London, 1982, pp. 149-176.
4. **van Dijk, T. 79:** Relevance Assignment in Discourse Comprehension. In: Discourse Processes, Vol. 2, No. 2, 1979, pp. 113-126.
5. **Dorfmueller-Karpusa, K. / Dorfmueller, T. 85:** Some Interdisciplinary Remarks about Coherence. In: Sözer, E. (ed): Text Connexity, Text Coherence. Papiere zur Textlinguistik Band 49, Hamburg, 1985, pp. 555-566.
6. **Fum, D. / Guida, G. / Tasso, C. 82:** Forward and Backward Reasoning in Automatic Abstracting. In: Proc. 9th Int. Joint. Conf. on Artificial Intelligence, 1982, pp. 840-844.
7. **Hahn, U. 87:** Lexikalisch verteiltes Text-Parsing. Eine objekt-orientierte Spezifikation eines Wortexpertensystems auf der Grundlage eines Aktorensystems. Dissertation, Konstanz, 1987.
8. **Hahn, U. 90:** Topic Parsing Accounting for Text Macro Structures in Full-Text. In: Information Processing & Management, Special Issue on Natural Language Processing and Information Retrieval. Vol. 26, No. 1, 1990, pp. 135-170.
9. **Hammwöhner, R. 90:** Automatischer Aufbau von Hypertext-Basen aus deskriptiv expositorischen Texten — Ein Hypertext-Modell für das Information-Retrieval. Konstanz, 1990, Dissertation.
10. **Hatakeyama, K. / Petöfi, J. / Sözer, E. 85:** Text, Connexity, Coherence. In: Sözer, Emel (ed): Text Connexity, Text Coherence. Papiere zur Textlinguistik Band 49, Hamburg, 1985, pp. 36-105.
11. **Heydrich, W. 89:** Relevant Objects and Situations. In: Conte, M.E. / Petöfi, J.S. / Sözer, E.: Text and Discourse Connectedness. Amsterdam / Philadelphia, 1989, pp. 413-425.
12. **Hobbs, J.R. 85:** On the Coherence and Structure of Discourse. Stanford University, Report CSLI-85-37, 1985.
13. **Horányi, ö. 85:** Discoherent Remarks on Coherence. In: Sözer, Emel (ed): Text Connexity, Text Coherence. Papiere zur Textlinguistik Band 49, Hamburg, 1985, pp. 567-594.
14. **Hutchins, J. 87:** Summarization : Some Problems and Methods. In: K. P. Jones (ed.): Informatics 9, Proc. of Informatics 9, King's College, Cambridge, 26-27 March 1987, London, 1987, pp. 151-173.
15. **Janoš, J. 79:** Theory of Functional Sentence Perspective and its Application for the Purposes of Automatic Extracting. Information Processing and Management, Vol. 15, No. 1, 1979, pp. 19-25. -
16. **Kieras, D.E. 82:** A Model of Reader Strategy for Abstracting Main Ideas from Simple Technical Prose. In: Text, Vol. 2, No. 1-3, 1982, pp. 47-31.
17. **Liebsch, H. / Werchosch, F.-U. 88:** Zur Integration von grafisch-figürlichen und lexisch-grammatischen Mitteln in der schriftlichen sprachlichen Kommunikation. In: Potsdamer

Forschungen der Pädagogischen Hochschule "Karl Liebknecht" Potsdam, Reihe A, Vol. 89, 1988. pp. 157-165.

18. **Mann, W.C. /Thompson, S.A. 88:** Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. In: Text, Vol. 8, No. 3, 1988, pp. 243-281.
19. **McKnight, C. / Richardson, J. / Dillon, A. 89:** Journal Articles as Learning Resource: What can Hypertext Offer. In: Jonassen, D. (ed): Proceedings NATO Advanced Research Workshop "Designing Hypertext/ Hypermedia for Learning". Rottenburg, FRG, 3.-7. Juli 1989. erscheint im Frühjahr 1990.
20. **Minsky, M. 75:** A Framework for Representing Knowledge. In: P.H. Winston (ed): The Psychology of Computer Vision, New York: McGraw-Hill, pp. 211-277.
21. **Reddig, C.84:** Word Expert Parsing: ein Überblick. In: Rollinger, C.R. (ed): Probleme des (Text-) Verstehens. Ansätze der künstlichen Intelligenz. Tübingen, 1984, pp. 77-88.
22. **Reimer, U. 86:** A Representation Construct for Roles. TOPIC-Report No. 16, Konstanz, 1986.
23. **Reimer, U. 89:** FRM: Ein Frame-Repräsentationsmodell und seine formale Semantik. Berlin, Heidelberg, 1989.
24. **Reimer, U. 90:** Automatic Acquisition of Terminological Knowledge from Texts. Erscheint in: Proc. ECAI 90.
25. **Reimer, U. / Hahn, U. 88:** Text Condensation as Knowledge Base Abstraction. In: Proceedings—The Fourth IEEE Conference on Artificial Intelligence Application, San Diego, California, Washington, D.C.: Comp. Soc. of the IEEE, 1988.
26. **Tait, J.I. 85:** Generating Summaries using a Scriptbased Language Analyzer. In: Steels, L. / Campbell, U.A. (eds): Progress in Artificial Intelligence, Chichester, 1985, pp. 312-318.
27. **Thiel, U. 90:** Konversationale graphische Interaktion mit Informationssystemen: Ein sprechakttheoretischer Ansatz. Konstanz, 1990, Dissertation.
28. **van de Velde, R. 85:** Inferences and Coherence in Text Interpretation. In: Sözer, E. (ed): Text Connexity, Text Coherence. Papiere zur Textlinguistik Band 49, Hamburg, 1985, pp. 261-298.