

Ein Kompositionalitätsprinzip für numerische Textsemantiken

Zusammenfassung: Der Beitrag beschreibt eine Variante des Kompositionalitätsprinzips der Bedeutung als Grundprinzip für die numerische Analyse unsystematischer Sinnrelationen komplexer Zeichen, das über das Phänomen der perspektivischen Interpretation hinaus gebrauchsssemantische Bedeutungsaspekte berücksichtigt. Ziel ist es, ein theoretisches Fundament für korpusanalytische Ansätze in der Semantik, die oftmals die linguistische Interpretierbarkeit ihrer Analyseergebnisse vermissen lassen, zu umreißen. Die Spezifikation des Kompositionalitätsprinzips erfolgt unter Rekurs auf das Modell eines hierarchisch geordneten Constraint-Satisfaction-Prozesses. Hiermit ist das längerfristige Ziel verbunden, das Problem einer defizitären numerischen Textrepräsentation sowie die mangelnde Integration von propositionaler und strukturaler bzw. korpusanalytischer Semantik anzugehen. Die Erörterungen dieses Beitrags sind primär konzeptioneller Natur; sie betreffen die Konzeption einer numerischen Textsemantik zur Vermeidung von Defiziten bestehender Ansätze.

1 Einleitung

Die Repräsentation natürlichsprachlicher Texte mit Hilfe numerischer Textrepräsentationsmodelle ist von entscheidender Bedeutung für Theorie und Praxis der automatischen Textanalyse, d.h. für die quantitative Textlinguistik ebenso wie für das anwendungsorientierte *text mining* (Mehler, 2002). Neuere Ansätze zur Simulation von Textverstehensprozessen und zur Quantifizierung der Kohärenz textueller Systeme verweisen darüber hinaus auf die zunehmende Bedeutung dieser Modelle im Bereich der kognitiven Linguistik (Kintsch, 1998). Textrepräsentationsmodelle spezifizieren in ihrer Gesamtheit, als welche Art mathematischer Entität Texte repräsentiert werden (als Mengensysteme, azyklische Grafen, Verteilungen, stochastische Prozesse, etc.).

Ein prominentes Beispiel für numerische Textrepräsentationsmodelle bilden *semantische Räume*, die in der kognitiven Linguistik (Foltz, 1996; Kintsch, 1998; Landauer/Dumais, 1997) und Computerlinguistik (Rieger, 1989; Schütze, 1998; Mehler, 2000) ebenso Anwendung finden, wie im Bereich des Information Retrievals (Deerwester et al., 1990). Die Grundidee semantischer Räume besteht darin, eine geometrische Interpretation von gebrauchsssemantischen Aspekten der strukturellen Bedeutung zu liefern: Je ähnlicher beispielsweise zwei Wörter im zugrunde liegenden Referenzkorpus gebraucht werden, desto näher sollen ihre Repräsentationen in Form so genannter Bedeutungspunkte (d.h. Vektoren) im semantischen Raum beieinander liegen. Als Beleg für die Güte semantischer Räume, deren Berechnung weitgehend automatisiert erfolgt, gilt

ihre Eignung, Nachbarschaftsrelationen auch solcher Zeichen abzubilden, die keine gemeinsamen Konstituenten besitzen oder niemals kookkurrieren. Ferner verzichten sie auf semantische Atome und verwenden stattdessen jene Wörter zur Beschreibung von Zeichen verschiedener Ebenen sprachlicher Resolution, deren Gebrauchsregularitäten im Referenzkorpus beobachtet werden. In diesem Sinne zielen semantische Räume auf die Abbildung gebrauchssemantischer Aspekte der strukturellen Bedeutung, die korpusanalytisch rekonstruierbar und durch Umgebungen von Bedeutungspunkten repräsentierbar sind.

Als zentrales Defizit dieses Formats erweist sich die unstrukturierte Repräsentation komplexer Zeichen. In der *latent semantic analysis* (LSA) werden beispielsweise Texte (unabhängig von ihrer Länge und Struktur) genauso wie Wörter je durch *einen* Bedeutungspunkt repräsentiert. Ferner bleiben semantische Räume auf den nicht-propositionalen Bedeutungsaspekt, genauer: auf die Abbildung unsystematischer Sinnrelationen, beschränkt. Unsystematisch bedeutet, dass Nachbarschaftsbeziehungen von Bedeutungspunkten nicht nach sinnrelationalen Kategorien (*Synonymie*, *Antonymie*, etc.) typisiert werden. Dies wirft die Frage nach der Integration von propositionaler und korpusanalytischer, struktureller Semantik auf, die in der Literatur weitgehend unbeachtet bleibt (siehe aber Ruge, 1995).

Im Rahmen der kognitiven Linguistik diskutiert nun Kintsch (2001) die Notwendigkeit einer Erweiterung des Repräsentationsformats „semantischer Raum“ zur adäquaten Erfassung von Sinnrelationen oberhalb der Wortebene, wobei er die Textebene explizit ausspart. Kintsch verfolgt das Ziel, Operationen auf semantischen Räumen einzuführen, welche die Kontextvariabilität lexikalischer Einheiten im Rahmen von Prädikat-Argument-Strukturen abzubilden erlauben. Der Begriff der Kontextvariabilität bezieht sich darauf, dass Wörter in immer neuen, nicht lexikalisierten, sprachlich manifestierten Kontexten ihre Bedeutung (mit welcher Resolution auch immer) modifizieren.¹ Um nun die Similaritätsrelationen von Prädikat-Argument-Strukturen $P(A)$ (als das Resultat der propositionalsemantischen Übersetzung einfacher Behauptungssätze) abzubilden, schlägt Kintsch eine „Prädikationsoperation“ $p_k: S^{k+2} \rightarrow S$ vor (S ist die Menge der Bedeutungspunkte), welche den Bedeutungspunkt $B(P(A)) \in S$ als das gewichtete Mittel der Bedeutungspunkte $B(A)$, $B(P)$ der lexikalischen Einheiten A , P und jener k unmittelbaren Nachbarn von $B(P)$ im semantischen Raum bestimmt, die den geringsten Abstand zu $B(A)$ aufweisen. Mit anderen Worten: Die Repräsentation der Bedeutung der Prädikat-Argument-Struktur $P(A)$ ist hier eine Funktion der Repräsentation von A und jener von P , und zwar unter dem Aspekt ihrer Relevanz für A .

¹ Dieses Phänomen wird im Rahmen der Diskussion um die Gültigkeit des Kompositionalitätsprinzips u. a. am Beispiel von NN- und AN-Konstruktionen diskutiert. Siehe hierzu Osherson/Smith (1981) und Kamp/Partee (1995) aber auch Lahav (1989), der diese Variabilität als ein Argument wider die Kompositionalität natürlicher Sprachen wertet.

Durch die rekursive Anwendung dieser Operation gelangt Kintsch zur Abbildung von Prädikat-Argument-Strukturen der Form $(P(A_1))(A_2)$ als Korrelate einfacher SVO-Sätze. Diese und andere Ansätze zur Abbildung von Phänomenen der Kontextvariabilität stehen einer strikten Anwendung des klassischen Kompositionalitätsprinzips (KP; siehe Janssen, 1997; Partee, 1984) entgegen: Zwar ist die Bedeutung von $P(A)$ eine Funktion der Bedeutungen der Konstituenten A , P und der Art ihrer syntaktischen Kombination, jedoch lässt diese Kennzeichnung die aspektabhängige Modifikation der Bedeutung von P unter der Perspektive des variablen Arguments A unbestimmt: Welcher Bedeutungsaspekt von P tatsächlich in die Superisation von $P(A)$ eingeht, variiert mit dem Argument A . Ferner muss die perspektivische Interpretation von P als Argument der Bedeutung von $P(A)$ nicht allein von der Bedeutung von A abhängen. Vielmehr kann bereits der Vortext jener Textstelle, an der $P(A)$ beobachtet wird, eine solche Selektionswirkung entfalten. Die Kognitionspsychologie spricht in diesem Zusammenhang von dem Phänomen des *Textprimings* (Sharkey/Sharkey, 1992), das sich darin zeigt, dass der Vortext einer Textstelle kontextadäquate Bedeutungsaspekte verstärkt, die Aktivierung kontextinadäquater Aspekte hingegen hemmt.

Der Konferenzbeitrag schlägt eine Variante des Kompositionalitätsprinzips der Bedeutung als ein Grundprinzip für numerische Ansätze zur Modellierung unsystematischer Sinnrelationen von Zeichen unterschiedlicher Ebenen sprachlicher Resolution vor, das über das Phänomen der perspektivischen Interpretation (bzw. des Textprimings) hinaus gebrauchsemantische Aspekte der Bedeutungskonstitution berücksichtigt. Die Relevanz dieses Vorschlags resultiert in erster Linie aus dem Versuch einer linguistischen, theoretischen Fundierung korpusanalytischer Ansätze in der Semantik, die oftmals die linguistische Interpretierbarkeit ihrer Analyseergebnisse vermissen lassen. Die Spezifikation des vorzuschlagenden „kontextsensitiven“ Kompositionalitätsprinzips erfolgt unter Rekurs auf das Modell eines hierarchisch geordneten *Constraint-Satisfaction*-Prozesses, und zwar unter Abgrenzung von Versuchen, die Kohärenz textueller Systeme als „ungeordneten“ *parallel constraint satisfaction process* zu modellieren (wie bei Thagard, 2000). Mit diesem Ansatz werden zwei zentrale Problempunkte semantischer Räume angegangen: das Ungenügen unstrukturierter Repräsentationen komplexer Zeichen sowie die mangelnde Integration von propositionaler und korpusanalytisch fundierter, struktureller Semantik.

Das einzuführende Prinzip wird als *semiotisches Kompositionalitätsprinzip* (SKP) bezeichnet, da es nicht (primär) den propositionalen, sondern strukturellen, gebrauchsemantischen Bedeutungsbegriff thematisiert. Es dient der regelgeleiteten Extension numerischer Textrepräsentationsmodelle auf den Bereich der strukturellen Semantik natürlichsprachlicher Texte und also der Entwicklung einer numerischen Textsemantik. In Abschnitt (2) wird hierzu das SKP informell umrissen und in Abschnitt (3) mittels *hierarchischer Constraint-Satisfaction-Probleme* genauer skizziert. Abschnitt (4) exemplifiziert das vorliegende Modell, während Abschnitt (5) einen Ausblick gibt.

2 Latente, semiotische Kompositionalität

Das SKP hat folgende vorläufige Gestalt:

Die Bedeutung einer sprachlichen Einheit y ist eine Relation über seinen Gebrauchsregularitäten (falls y lexikalisiert ist), den Kontexten seiner Vorkommen (als Systeme syntaktischer Dependenz- und textueller Kohärenzrelationen, an denen y teilhat), den Bedeutungen seiner Komponenten und der Art ihrer Kombination sowie über beschriebenen Situationen. Die Interpretation von y in einem gegebenen Kontext ist jene Situation, die y unter den Bedingungen konkreter Instanziierungen dieser Parameter beschreibt.

Den Hintergrund dieser „Definition“ bildet der relationale Bedeutungsbegriff der Situationssemantik (Barwise/Perry, 1983), der auf zweierlei Weisen erweitert wird: durch Rekurs auf Kohärenzrelationen als Bestimmungsfaktoren der kontextsensitiven Interpretation sprachlicher Einheiten sowie durch Rekurs auf den Begriff der Gebrauchsregularität, demzufolge die kontextspezifische Interpretation elementarer Zeichen jene Bedingungen verändern kann, die ihrem Gebrauch zugrunde liegen. Diese Kennzeichnung ist jedoch mit fundamentalen Problemen verbunden, was vor allem die mangelnde Aufzählbarkeit von Interpretationen, Interpretationskontexten und Gebrauchsregularitäten betrifft, die je für sich informationell ungewisse, dynamische Entitäten darstellen und also nicht qua Lexikonreferenz selektierbar sind. Um diese Dynamik zu erfassen, ist das SKP unter Rekurs auf prozedurale Konstituenten als Modelle jener kognitiven Prozesse zu reformulieren, die letztere Entitäten konstituieren:

Die Bedeutung einer sprachlichen Einheit y ist eine Prozedur P , die ihre Interpretationen in Abhängigkeit von drei Parametern generiert: ihren Gebrauchsregularitäten (siehe oben), den Kontexten ihres Vorkommens (siehe oben), den Bedeutungen ihrer Komponenten und der Art ihrer Kombination. Dabei wird verlangt, dass P semiotisch interpretierbar ist. Eine Prozedur heißt semiotisch interpretierbar, wenn nicht nur ihre Input- und Outputgrößen Modellfunktion besitzen, sondern auch P als Ganzes.

Das Kriterium der semiotischen Interpretierbarkeit bedingt die Suche nach Modellen, die nicht nur „gute Ergebnisse“ produzieren, sondern aufgrund ihrer Funktionsweise semiotisch interpretierbar sind und also (kognitiven/sozialen) syntaktischen, semantischen oder pragmatischen Prozessen entsprechen. Im Folgenden wird dieses Kriterium durch Rekurs auf das Konstruktions-Integrations-Modell (KI-Modell) von Kintsch (1998) berücksichtigt. Analysiert man diese nach wie vor vorläufige Definition, so sind (wie schon im Falle des KPs – Janssen, 1997) zahlreiche Parameter auszumachen. Dies betrifft u.a. den thematisierten Bedeutungs- und Kontextaspekt, den involvierten Begriff der sprachlichen Einheit, ihrer Komponenten und der Art ihrer Kombination sowie den Begriff der Gebrauchsregularität. Das SKP ist in der vorliegenden Form somit inhärent unterbestimmt. Im Folgenden wird diese Mehrdeutigkeit nicht vollstän-

dig beseitigt. Vor dem Hintergrund obiger Einleitung steht jedoch die Beschränkung auf den strukturellen Bedeutungsaspekt fest, demzufolge ein geeignet modifiziertes KP dem Umstand Rechnung zu tragen hat, dass die unsystematischen Sinnrelationen komplexer sprachlicher Einheiten systematisch mit den Sinnrelationen ihrer Teile zusammenhängen.

3 Constraint-Satisfaction mit hierarchischer Ordnung

Ohne das SKP für verteilte, numerische Semantiken vollständig zu formalisieren, sollen nun grundlegende Konstituenten dieses Prinzips erläutert werden. Hierzu wird das Prinzip zur Charakterisierung eines Constraint-Satisfaction-Prozesses der Textinterpretation herangezogen, der durch die hierarchische Ordnung von Textkomponenten überlagert wird. Den strukturellen Kern dieses Textmodells bilden die so genannte Integrationshierarchie des Inputtexts, die (syntaktischen Dependenz-) und textuellen Kohärenzrelationen seiner Komponenten sowie der operative Rezeptionskontext. Die Integrationshierarchie reflektiert den Zusammenhang, dass Texte im Erfolgsfalle ausgehend von elementaren über komplexe Komponenten bis hinauf zum Textganzen interpretiert werden. Sie dient zur Ordnung sämtlicher Prozesse der diskursspezifischen Interpretation von informationell ungewissen (mehrdeutigen, elliptischen, vagen) Komponenten, die auf Einheiten unterschiedlicher Ebenen textueller und kontextueller Resolution rekurrieren können. Aus prozeduraler Sicht ist dieses Textmodell wie folgt angelegt: Der Inputtext konstituiert kraft Integrationshierarchie und Kohärenzrelationen (als informationell ungewisse Interpretationsrestriktionen) ein *hierarchisches Constraint-Satisfaction-Problem* (HCSP), dessen Lösung darin besteht, seine Komponenten entlang der Integrationshierarchie bis hinauf zum Textganzen so zu interpretieren, dass möglichst viele der möglichst wichtigen Kohärenzrelationen, welche die Komponenten auf den jeweiligen Ebenen eingehen, berücksichtigt werden.

In Abhängigkeit von den operativen Kohärenzrelationen und der Art von Textur, die sie aufspannen, kann dieser Interpretationsprozess im Extremfall strikt kompositional verlaufen, ohne durch Kohärenzrelationen, welche die Integrationshierarchie überschreiten, konterkariert zu werden. Im umgekehrten Extremfall erfolgt er streng nach dem Kontextprinzip (Janssen, 1997), und zwar unter prinzipieller Umgehung der Integrationshierarchie. In diesem Sinne deckt das SKP ein Kontinuum ab, das von der strikten Kompositionalität (wie sie die logische Semantik kennzeichnet) bis hin zur strikten Kontextualität (die für konnektionistische Ansätze nach dem Prinzip des *parallel constraint satisfaction* charakteristisch ist) reicht. Welches Verhältnis von Kompositionalität und Kontextualität in Texten tatsächlich realisiert wird, dürfte u.a. mit dem zugrunde liegenden Texttyp variieren.

Zur leichteren Einordnung der folgenden Erörterungen seien zwei Unterscheidungen vorausgeschickt: Ein Diskurs wird als ein System bestehend aus einem Text und einem situativen/kognitiven/sozialen Rezeptionskontext aufgefasst (Produktionsprozesse bleiben der Einfachheit halber ausgespart). Dies hat zur Folge, dass Kohärenzrelationen als Relationen (kon-)textueller Einheiten definierbar werden, die u.a. dazu dienen, Texte in ihre Rezeptionskontexte einzubetten. Diesem Ansatz entspricht die situationssemantische Unterscheidung von Bedeutung und Interpretation: Sprachliche Einheiten werden kontextsensitiv interpretiert, wozu sie eine sprachsystematische Bedeutung beisteuern, die unter Abstraktion individueller Rezeptions-/Produktionskontexte beschrieben wird. In diesem Sinne kann von der sprachlichen Textbedeutung und der diskursspezifischen Textinterpretation gesprochen werden. Im Folgenden werden strukturelle und prozedurale Aspekte von HCSPen und ihre schematische Lösung erläutert (wobei die Erläuterungen auf monologische, schriftsprachliche Texte beschränkt bleiben). Wir beginnen mit den strukturellen Bausteinen:

(1) *Integrationshierarchie*: In HCSPen interagieren Textkomponenten entlang von (syntaktischen Dependenz- und textuellen) Kohärenzrelationen zur Eingrenzung ihrer diskursspezifischen Interpretation, auf deren Basis sie an der Integration komplexerer Komponenten teilhaben. Dieser Prozess erfolgt nicht parallel unter gleichberechtigter Interaktion aller Komponenten. Vielmehr erhält er durch die Integrationshierarchie eine Ordnung, derzufolge Texte eine hierarchische Grundstruktur (mit dem Text als Wurzel und seinen elementaren Komponenten als Blätter) besitzen. Es ist diese Kernstruktur, die durch Kohärenzrelationen überschritten wird, indem sie Textkomponenten mit jenen ko- und kontextuellen Einheiten verbinden, die ihre Interpretation restringieren, und also poly-hierarchische, netzwerkartige Strukturen erzeugen können. Im Folgenden wird die Integrationshierarchie eines Texts mit seinem Strukturbaum gemäß der *rhetorical structure theory* (Mann/Thompson, 1988), bisweilen aber auch mit seiner oberflächenstrukturellen Unterteilung in Sektionen, Absätze, Sätze, etc. im Sinne der *ordered hierarchy of content objects* (Renear et al., 1996) identifiziert.

(2) *Kohärenztypen*: In HCSPen werden Interpretationsrestriktionen analog zur Unterscheidung von *Tupeln*, *Relationen*, *Relationenformaten* und *intra- bzw. interrelationalen Abhängigkeiten* des Relationenkalküls spezifiziert – als Teil des Kompositums *Kohärenzrelation* wird *Relation* dabei nicht mathematisch, sondern linguistisch gebraucht: *Kohärenzrelationen* (*Tupel* im mathematischen Sinne) von Diskurskomponenten konstituieren *Kontexte* als Mengen von Kohärenzrelationen (und also Grafen induzierende mathematische *Relationen*). Diese Kontexte definieren komplexe Beschränkungen zur Interpretation von Textkomponenten, und zwar gemäß den Semantiken jener *Kohärenztypen* (als Entsprechungen von *Relationenformaten*), welche die Kohärenzrelationen instanzieren. Diese Typen spezifizieren, auf welche Weisen Textkomponenten mit ihren Kontexten zusammenhängen können und wie diese Verknüpfungen zu interpretieren sind. Sie variieren hinsichtlich ihrer Signatur (d.h. nach

der (variablen) Zahl und den Rollen ihrer obligaten/optionalen Argumente), der Weise, wie ihre Instanzen (Kohärenzrelationen) die Interpretationen ihrer Argumente restringieren, sowie im Hinblick auf die sprachlichen Muster, durch die sie oberflächenstrukturell realisiert werden. Ein Kohärenztyp kann somit als ein Tripel

$$\langle (A_i)_{i \in I}, (\text{real}_i)_{i \in I}, K \rangle, \quad (1)$$

dargestellt werden, das aus einer Konstruktionsprozedur K (siehe unten), einer Familie von Attributen $(A_i)_{i \in I}$ zur Spezifikation von Argumentrollen sowie aus einer Familie $(\text{real}_i)_{i \in I}$ von Funktionen real_i zur Abbildung des i -ten Attributs A_i auf eine Repräsentation seiner Realisierungsmuster besteht, die als probabilistische Systeme von lexikogrammatistischen Auswahlmöglichkeiten zur oberflächenstrukturellen Indikation von Argumentrollen beschrieben werden können. So umfasst z.B. die anaphorische Referenz als Kohärenztyp im einfachsten Fall die Rolle der Anapher und des Antezedens, wobei erstere vorzugsweise durch Pronomina und letztere durch Nomina realisiert wird. Im Gegensatz hierzu umfasst der Typ der Elaboration nach Mann/Thompson (1988) einen Nukleus in der Rolle der elaborierten und einen Satelliten in der Rolle der elaborierenden Textspanne, wobei beide Rollen zumindest von Einheiten der Phrasenebene realisiert werden. Ein System von Kohärenzrelationen beschreibt beispielsweise Martin (1992). Die sicherlich probabilistische Grammatik der zugehörigen Realisierungsmuster bleibt hier aus Platzgründen ausgespart.

(3) Eine *Konstruktionsprozedur* (K) ist ein prozedurales Modell (Sanders/Spooren, 2001) der Semantik eines Kohärenztyps. Sie legt fest, wie die Interpretation einer Textkomponente durch jene Komponente eingeschränkt wird, mit der sie eine Kohärenzrelation des betrachteten Typs bildet. Der Rekurs auf Konstruktionsprozeduren reflektiert den Umstand, dass HCSPE nicht nur Integrationsprozesse, sondern auch Prozesse der Konstruktion von Zwischenrepräsentationen und Interpretationsalternativen in Analogie zum KI-Modell von Kintsch (1998) berücksichtigen, und zwar unter Abgrenzung von Thagards (2000) Modell der Kohärenz als mengentheoretisches Separierungsproblem.

Zur Konkretisierung von Konstruktionsprozeduren rekuriert der vorliegende Ansatz auf ein numerisches Format zur Repräsentation der *strukturellen Bedeutungen* und *kontextsensitiven Interpretationen* sprachlicher Einheiten, das im Folgenden als *semiotischer Raum* bezeichnet wird. Anders als in semantischen Räumen, in denen Wörter und Texte undifferenziert repräsentiert werden, dienen semiotische Räume dazu, Zeichen auf Hypergraphen ihrer Punkte abzubilden, die ihre Bedeutungen bzw. Interpretationen *und* Strukturen modellieren. Sie erweitern in dieser vorläufigen Ausbaustufe semantische Räume durch Hinzunahme des Kriteriums der Struktursensitivität von Repräsentationen. Vor diesem Hintergrund können Konstruktionsprozeduren als abstrakte Modelle kognitiver Prozesse durch Operationen auf semiotischen Räumen als Datenstrukturen zur Abbildung mentaler Textrepräsentationen konkretisiert werden (sie-

he unten), und zwar in Einklang mit der Unterscheidung von prozeduraler, algorithmischer und implementationsnaher Modellierung (Bierwisch, 1988).

(4) *Kotext, Kontext und Textur*: Kohärenzrelationen werden als diskurskontextspezifische Instanzen von Kohärenztypen aufgefasst und (im Falle ungerichteter Relationen) als Mengen bzw. (im Falle gerichteter Relationen) als Tupel von (Mengen von) Diskurskomponenten repräsentiert. Sie induzieren Beschränkungen über den Interpretationen ihrer Argumente, und zwar gemäß der Semantik des zugehörigen Kohärenztyps und der Rollen, welche die Argumente einnehmen. Vor diesem Hintergrund kann der *Kontext* einer Textkomponente definiert werden als das System jener Kohärenzrelationen, an denen sie teilhat. Er legt fest, unter der Perspektive welcher Interpretationsrestriktionen (als das Ergebnis der Wirksamkeit) welcher Kohärenztypen (kraft der Verknüpfung mit welchen Komponenten) die fokale Komponente zu interpretieren ist. Aus Vereinfachungsgründen bleiben Diskurskomponenten im Folgenden auf textuelle Einheiten beschränkt. Der Kontext einer Textkomponente umfasst also Einheiten desselben Texts oder textuelle Manifestationen des situativen/kognitiven/sozialen Rezeptionskontexts. Bleibt der Kontext auf Komponenten desselben Texts beschränkt, so wird genauer von *Kotext* gesprochen. Schließlich kann die *Textur* eines Diskurses definiert werden als die Vereinigung der Kontexte seiner Komponenten und der Integrationshierarchie des zugehörigen Texts, dessen Textur wiederum über die Integrationshierarchie hinaus auf solche Kohärenzrelationen beschränkt bleibt, die Textkomponenten eingehen. Während also Kohärenztypen sprachsystematische Größen bilden, sind Kohärenzrelationen, Ko(n)texte und Texturen diskurs- bzw. textspezifische Einheiten.

(5) *Semiotische Räume*: Kintsch (1998) rekurriert auf semantische Räume von Landauer/Dumais (1997), um Verknüpfungen von Knoten in Netzwerken propositionaler Textrepräsentationen zu gewichten. Semiotische Räume rekurrieren umgekehrt auf textuelle Kohärenz- und syntaktische Dependenzrelationen, um Kanten von Hypergraphen vektorieller Zeichenrepräsentationen zu typisieren. Dieses Vorgehen steht in Einklang mit dem Ansatz von Kintsch (2001) zur Erweiterung der Menge von Operationen auf semantischen Räumen, weicht jedoch in zweierlei Hinsicht von ihm ab: Erstens werden textuelle Einheiten nicht ausgeschlossen, sondern ebenso numerisch analysiert wie Wörter und einfache Prädikat-Argument-Strukturen. Zweitens erfolgt diese Analyse unter der Kontrolle des SKPs. Die negative Einschätzung einer struktursensitiven numerischen Analyse der Bedeutungen von Texten bei Kintsch ist unseres Erachtens dem Ungenügen des verwendeten Repräsentationsformats in Form des semantischen Raums geschuldet. Demgegenüber differenzieren semiotische Räume elementare und komplexe Einheiten dadurch, dass erstere auf Punkte, letztere hingegen auf Hypergraphen solcher Punkte abgebildet werden. Dies gilt für sprachsystematische Bedeutungen ebenso wie für diskurspezifisch zu konstruierende Interpretationen. Ausgehend von der fokalen Textkomponente y können nun semiotische Räume durch

Rekurs auf das Schema $\langle X, G_1, G_2 \rangle$ konkretisiert werden, in dem X als Teilmenge des semiotischen Raums die (informationell ungewisse) Bedeutung bzw. Interpretation von y repräsentiert, während die beiden Grafen G_1 und G_2 , deren Knoten auf Teilmengen von Punkten des semiotischen Raums beschränkt sind, die Integrationshierarchie bzw. den jeweiligen Diskurskontext von y modellieren. Vor diesem Hintergrund sind elementare und komplexe Textkomponenten wie folgt zu unterscheiden:

- Die *Bedeutung* $B(y)$ der *elementaren Komponente* y des Texts x (d.h. die sprachsystematische Bedeutung eines Blatts seiner Integrationshierarchie) wird als Tripel $B(y) = \langle Y, G_1, G_2 \rangle$ modelliert, wobei Y eine Teilmenge des semiotischen Raums S ist, welche die alternativen Bedeutungen der möglicherweise mehrdeutigen Komponente y repräsentiert. Ferner sind $G_1 = G_2 = \langle \{y\}, \{\} \rangle$ zwei Grafen mit leeren Kantenmengen, die dem Sachverhalt Ausdruck verleihen, dass elementare Komponenten keine Integrationsstruktur besitzen und die Bestimmung ihrer sprachsystematischen Bedeutung unabhängig vom Kontext erfolgt. Demgegenüber wird die diskurspezifische *Interpretation* $I(y, |x|)$ von y als Komponente des Diskurses $|x|$ als Tripel $I(y, |x|) = \langle Y', G_1, Co(y, |x|) \rangle$ modelliert, in dem Y' abermals eine Teilmenge des semiotischen Raums ist, $G_1 = \langle \{y\}, \{\} \rangle$ der bereits erwähnte Graf zur Repräsentation der Strukturlosigkeit von y und $Co(y, |x|)$ ein Hypergraf, welcher den Kontext von y in $|x|$ repräsentiert. Die Divergenz von Bedeutung und Interpretation kommt in diesem Modell dadurch zum Ausdruck, dass die Mengen Y und Y' divergieren können, wobei eine geringere Kardinalität letzterer Menge den Umstand reflektiert, dass die kontextsensitive Interpretation von y spezifischer ist als ihre Bedeutung.
- *Komplexe Komponenten*: Die *Bedeutung* einer im Sinne der Integrationshierarchie $H(x)$ komplexen Komponente y wird rekursiv definiert:

$$B(y) = \langle Y, G_1, G_2 \rangle, \quad (2)$$

wobei Y eine Teilmenge des semiotischen Raums S ist, während $G_1 = \langle \{Y, Y_1, \dots, Y_n\}, E \rangle$ jenen Teil der Integrationshierarchie $H(x)$ modelliert, der durch y und seine unmittelbaren Komponenten y_1, \dots, y_n aufgespannt wird, wobei die Teilmengen Y_1, \dots, Y_n den Bedeutungsrepräsentationen $B(y_1), \dots, B(y_n)$ angehören. Darüber hinaus resultiert Y annahmegemäß aus der Anwendung einer so genannten *Integrationsoperation* O_1 auf die Mengen Y_1, \dots, Y_n . Der dritte Parameter G_2 ist wieder ein Graf mit leerer Kantenmenge. Die *Interpretation* von y als Komponente des betrachteten Diskurses $|x|$ wird demgegenüber wie folgt repräsentiert:

$$I(y, |x|) = \langle Y', G_1, Co(y, |x|) \rangle, \quad (3)$$

Y' ist wieder eine Teilmenge des semiotischen Raums und $G_1 = \langle \{Y', Y'_1, \dots, Y'_n\}, E' \rangle$ ein Graf, der jene Teilmengen Y'_1, \dots, Y'_n als Knoten umfasst, die den Interpre-

tationen $I(y_1, |x|), \dots, I(y_n, |x|)$ der unmittelbaren Konstituenten von y gemäß $H(x)$ angehören. Anders als zuvor jedoch beruht die Berechnung von Y' nicht allein auf den Interpretationen dieser Konstituenten, sondern zugleich auf dem Kontext $Co(y, |x|) = \langle V, D \rangle$ von y in $|x|$, der als ein Hypergraf modelliert wird, in dem sämtliche Kohärenzrelationen, an denen y teilhat, Kanten bilden und dessen Knotenmenge V aus der Vereinigung aller Knoten dieser Kanten resultiert. Vor diesem Hintergrund kann die Berechnung von Y' wie folgt skizziert werden: Zunächst wird eine Zwischenrepräsentation Y'' durch Anwendung der Integrationsoperation O_1 auf die Teilmengen Y'_1, \dots, Y'_n erzeugt. Im Anschluss hieran werden jene Konstruktionsoperationen O parallel auf Y'' angewandt, die (nach Formel 1) den prozeduralen Semantiken der Kohärenztypen der in $Co(y, |x|)$ wirksamen Kohärenzrelationen entsprechen. Kanten $e \in E$, die y mit nachgeordneten Komponenten z kataphorisch verknüpfen, welche zum aktuellen Verarbeitungszeitpunkt uninterpretiert sind, werden erst mit der Interpretation von z ausgewertet. Eine Komponente z heißt nachgeordnet in Bezug auf y , wenn es innerhalb der Integrationshierarchie einen Knoten gibt, der y in einem linken und z in einem rechten Teilbaum dominiert. In allen anderen Fällen gehen die Interpretationen der verknüpften Komponenten in den Prozess der Interpretation von y ein.

Letztere Schritte betonen die spezifische Charakteristik von HCSPen, die den Mechanismus der parallelen Auswertung von Beschränkungen mittels der Integrationshierarchie ordnen. Dabei können die Interpretationen der fokalen Komponente y und ihrer elementaren Konstituenten – bedingt durch die wirksamen Kohärenzrelationen – von ihren jeweiligen sprachsystematischen Bedeutungen abweichen. Und also hängt die Interpretation von y nicht nur von den Bedeutungen ihrer Teile und der Art ihrer Kombination ab, sondern zugleich von ihren Kohärenzrelationen, die über jenen Ausschnitt der Integrationshierarchie hinaus gehen können, den y dominiert. Die geometrische Entsprechung dieses Prozesses bilden Konstruktionsoperationen, die eine Verlagerung, Ausweitung oder Verkleinerung jener Region(en) des semiotischen Raums bewirken (können), welche diese systematische(n) Bedeutung(en) repräsentieren. Der Grad der Abweichung von sprachsystematischer Bedeutung und diskurspezifischer Interpretation ist dabei eine Funktion der kontextuellen Einbettung von y . Konstruktion und Integration sind kooperative Prozesse der Textinterpretation und werden als *prozedurale Konstituenten von HCSPen* wie folgt unterschieden:

(1) *Integration*: Die primäre Ordnung der Textinterpretation folgt der Integrationshierarchie: Der Interpretationsprozess beginnt mit den Blättern dieses Baums und endet im Erfolgsfall mit seiner Wurzel, dem Gesamttext. Im Falle intermediärer Komponenten sind es die im unmittelbar vorangehenden Integrationsschritt (bzw. Interpretationsschritt) integrierten (bzw. konstruierten) Bedeutungen (bzw. Interpretationen), die als Argumente in diesen Teilprozess eingehen.

(2) *Konstruktion*: Sobald die fokale Textkomponente auf der Basis der Interpretationen ihrer unmittelbaren Konstituenten integriert ist, setzt ein Konstruktionsprozess ein, der durch die Interpretationen jener Komponenten restringiert wird, welche den Kotext der fokalen Komponente aufspannen. Dieser Prozess entfaltet sich entlang der jeweils wirksamen Kohärenzrelationen, wobei sein konstruktives Moment u.a. darin zum Ausdruck kommt, dass er eine Teilmenge des semiotischen Raums als Interpretation der fokalen Komponente bestimmt, die sämtliche, im Einklang mit den operativen Interpretationsrestriktionen stehenden, alternativen Interpretationen umfasst. Anders als bei Kintsch sind es jedoch Kohärenzrelationen, die hier als logische Atome der Diskursorganisation dienen und den kontextsensitiven Konstruktionsprozess steuern. Da die elementaren Textkomponenten nicht integriert werden, bilden ihre sprachsystematischen Bedeutungen den initialen Input des Konstruktionsprozesses, wiederum unter den Restriktionen ihrer Kohärenzrelationen.

(3) *Abfolgerichtung*: Unter Berücksichtigung linearer, hierarchischer und netzwerkartiger Aspekte der Diskursorganisation erfolgt der Wechsel von Konstruktions- und Integrationsprozessen von links nach rechts und von unten (beginnend mit den Blättern der Integrationshierarchie) nach oben (bis hin zur Wurzel), wobei top-down operierende Prozesse des Kontextprimings und rechts-links gerichtete Revisionsprozesse hinzutreten. Top-down gerichtete Prozesse basieren auf solchen Kohärenzrelationen, die Textkomponenten mit Einheiten des situativen, kognitiven oder sozialen Kontexts verbinden, wie im Falle schematischer, generischer (Super-)Strukturen (Dijk/Kintsch, 1983; Martin, 1992). Revisionsprozesse basieren demgegenüber auf kataphorischen Verknüpfungen von Komponenten y_i mit nachgeordneten Komponenten y_j , deren zeitlich ebenfalls nachgeordnete Interpretation infolge der Auswertung ihres Kontexts eine Modifikation der zu diesem Zeitpunkt gültigen Interpretation von y_i (und jener Komponenten, die y_i in $H(x)$ dominieren) bedingen kann. Die Einleitung dieser Revision (bzw. Bestätigung) ist abermals an die Integrationshierarchie gebunden und kehrt ebenso vom Paradigma des *parallel constraint satisfaction* ab.

Konstruktionsoperationen sind algorithmische, auf semiotischen Räumen operierende Spezifikationen von Konstruktionsprozeduren als Modelle kognitiver Prozesse des Textverstehens. Ihre Konstruktivität liegt darin, Interpretationen nicht lexikonbasiert auszuwählen, sondern diskursspezifisch zu erzeugen. Um HCSPe als Modelle der Textinterpretation zu bilden, sind die Integrations- und Konstruktionsoperationen für alle berücksichtigten Kohärenztypen zu definieren. Beide Operationstypen entsprechen Teilprozessen der Textinterpretation: dem sich entlang der Integrationshierarchie entfaltenden Integrations- bzw. dem sich entlang von Kohärenzrelationen (als ungewisse Informationsbasen zur Bestimmung kontextadäquater Interpretationen) entfaltenden Konstruktionsprozess. Dieser Ansatz bewahrt die Architektur von Kintschs KI-Modell, auch wenn er auf das propositionale Repräsentationsformat verzichtet und eine andere Semantik von Integrations- und Konstruktionsprozessen vorsieht.

(4) *Gebrauchssemantik*: Gemäß der *weak contextual hypothesis* von Miller/Charles (1991) ist die semantische Ähnlichkeit von Wörtern durch die Ähnlichkeit ihrer kontextuellen Repräsentationen bedingt. Um den hiermit verbundenen Lernprozess als das Resultat von Textverstehensprozessen abzubilden, werden die kontextsensitiv interpretierten Textkomponenten zur inkrementellen Revision der Bedeutungsrepräsentationen ihrer elementaren Konstituenten herangezogen: Sei $B_i(a)$ die Bedeutungsrepräsentation von Wort a als das Ergebnis der Verarbeitung des Korpus $C = \{x_1, \dots, x_i\}$. Die inkrementelle Revision von $B_i(a)$ als das Resultat der Verarbeitung der Textsequenz x_{i+1}, \dots, x_{i+n} , unter den Bedingungen ihrer jeweiligen Kontexte $|x_{i+1}|, \dots, |x_{i+n}|$ kann dann wie folgt symbolisiert werden:

$$B_i(a) \xrightarrow{|x_{i+1}|} B_{i+1}(a) \xrightarrow{|x_{i+2}|} \dots \xrightarrow{|x_{i+n-1}|} B_{i+n-1}(a) \xrightarrow{|x_{i+n}|} B_{i+n}(a) \quad (4)$$

Der solcherart symbolisierte Prozess einer inkrementellen Konstitution struktureller Bedeutungen als das Ergebnis eines diskursseitigen, kontinuierlich bestätigten Interpretationsprozesses verweist auf einen weiteren Aspekt, unter dem das HCSP-Modell vom klassischen Modell rein synchroner Kompositionalität abweicht. Diese Abkehr ist unabdingbar, wenn (wie im Falle der strukturellen Bedeutung) solche Bedeutungsaspekte betrachtet werden, die durch informationelle Ungewissheit gekennzeichnet sind.

Vor diesem Hintergrund kann nun jenes *Optimierungsproblem* charakterisiert werden, das HCSPe induzieren:

(1) *Konsonante vs. dissonante Beschränkungen*: Ausgehend von einer fokalen Textkomponente y können Kohärenzrelationen, die ihren Kontext aufspannen, kooperieren oder konkurrieren. Betrachtet man z.B. Relationen des Typs der unsystematischen lexikalischen Kohäsion nach Halliday/Hasan (1976), so kann y mit Wörtern desselben Texts verknüpft sein, die konsonante oder dissonante Interpretationen von y unterstützen (wie im Falle mehrdeutiger Lexeme im Kontext von Wörtern, die ihre verschiedenen Lesarten unterstützen). Auf der Ebene der rhetorischen Kohärenz tritt ein analoger Fall auf, wenn dieselbe Textspanne auf der Basis von Hintergrundrelationen mit verschiedenen, einander widersprechenden Textspannen verknüpft ist. Wie sind solche Kohärenzrelationen im Zuge der Interpretation von y auszuwerten? In allgemeiner Lesart (und ohne Berücksichtigung der genauen Semantik der involvierten Kohärenztypen) kann die Antwort wie folgt umrissen werden: Konsonante Kohärenzrelationen kooperieren zum Zwecke der Eingrenzung jener Region des semiotischen Raums, in der die diskursspezifische Interpretation von y lokalisiert ist – im Extremfall wird hierdurch genau ein Punkt (bzw. Vektor) eingegrenzt. Dissonante Kohärenzrelationen bewirken das Gegenteil: Hier kann der Konstruktionsprozess eine größere, möglicherweise diskontinuierliche Region als potentiellen Ort der zu ermittelnden Interpretation abgrenzen. Offensichtlich können solche Relationen die unvollständige Lösung eines HCSPs bedingen, bei der die Interpretation einer Komponente unspezifisch weite

Teile des semiotischen Raums umschließt. Dieser Umstand folgt der Einsicht, dass Kohärenz eine graduierbare Eigenschaft von Textrepräsentationen ist.

(2) *Negative Beschränkungen*: Einen Sonderfall bilden negative Beschränkungen, d.h. Kohärenzrelationen (wie z.B. die Kontrastrelation), welche die Interpretation einer Komponente als etwas restringieren, was sie *nicht* ist. Wie sind solche Beschränkungen zu berücksichtigen? Ohne eine detaillierte Antwort zu geben, sei der Fall betrachtet, dass bereits eine vorläufige Interpretation von y existiert. Die Berücksichtigung einer negativen Beschränkung kann dann bedeuten, jene Teilregionen innerhalb dieser Interpretation zu bestimmen, die durch die Beschränkung nicht ausgeschlossen werden.

(3) *Optimierung*: Die strukturelle und prozedurale Konstitution von HCSPen induziert ein Optimierungsproblem, das folgendermaßen zu umschreiben ist: Ein HCSP zu lösen bedeutet, entlang der Integrationshierarchie des Inputtexts von links nach rechts, und von unten nach oben so komplexe Komponenten wie möglich zu integrieren und dabei auf jeder Stufe der Integrationshierarchie so viele Interpretationsrestriktionen wie möglich zu erfüllen, und zwar in der absteigenden Reihenfolge ihrer Wichtigkeit. Hierzu wird auf jeder Stufe der Integrationshierarchie eine Teilmenge des semiotischen Raums eingegrenzt, deren Kardinalität desto geringer ist, je spezifischer, informationell eindeutiger, konsonanter die zu berücksichtigenden Beschränkungen sind.

Abschließend kann nun unter weitgehender Berücksichtigung der vorangehenden Festlegungen das SKP formaler dargestellt werden: Sei y eine Komponente von x mit den unmittelbaren Konstituenten y_1, \dots, y_n in $H(x)$; symbolisch: $y = f(y_1, \dots, y_n)$. Dann ist der Mechanismus des SKPs in Bezug auf y durch folgende „situationsemantische“ Notation charakterisierbar:

$$\begin{aligned} Co(y, |x|) \| y \| D &= Co(y, |x|) \| f(y_1, \dots, y_n) \| D \\ &= Co(y, |x|) (\| f \|) [Co(y_1, |x|) \| y_1 \| D_1, \dots, Co(y_n, |x|) \| y_n \| D_n] D \end{aligned} \quad (5)$$

Dabei ist $Co(y, |x|)$ der Kontext von y im betrachteten Diskurs $|x|$, D die Interpretation von y gemäß Formel (3) und $\| y \|$ das Ergebnis der Integration von y auf der Basis der Interpretationen seiner unmittelbaren Konstituenten, die schließlich auf die Interpretationen ihrer elementarer Komponenten zurückgeführt werden. Dieser Formel gemäß ist die Interpretation D von y im Kontext $Co(y, |x|)$ gleich der Interpretation jenes Resultats der Integration seiner kontextsensitiv interpretierten Teile, die aus der Anwendung der Integrationsoperation $\| f \|$ resultiert. Auch wenn diese Formel (quasi in Analogie zum klassischen KP) auf einem Homomorphismus von Ausdrucks- und Inhaltsplan zu basieren scheint, ist es (von der relationalen Schreibweise abgesehen) der Kontext $Co(y, |x|)$ der Komponenten y_i , $i \in \{1, \dots, n\}$, der die Integrationshierarchie überschreiten kann und also einer strikten Anwendung des KPs entgegensteht.

4 Ein Analysebeispiel

Um das Konzept des HCSPs (und indirekt des SKPs) zu verdeutlichen, soll nun ein Textbeispiel und seine Analyse auf der Basis von vier Ansätzen betrachtet werden: des Vektorraummodells (Salton, 1989), der LSA (Landauer/Dumais, 1997), des Textraummodells (Mehler, 2000) und des HCSP-Modells. Das Textbeispiel lautet:

[Nukleus:] Alle Skulpturen werden restauriert. [Satellit:] Nur die Löwen bleiben im Park.

Die Wahl dieses Beispiels ist dadurch motiviert, dass Konstruktionen der Art von *stone lion* eine wesentliche Rolle in der Diskussion um das KP spielen (siehe Kamp/Partee, 1995) und prinzipiell in der von Kintsch (2001) erörterten Weise analysierbar sind. Ohne nun die vier Verfahren *en détail* zu erläutern, können die Grundprinzipien ihres textanalytischen Vorgehens wie folgt charakterisiert werden:

1. Das Vektorraummodell fasst den Text als eine Wortmenge auf, aus der Funktions- und andere Wörter gefiltert werden. Die verbleibenden Wörter (etwa *Skulptur*, *restaurieren*, *Löwe*, *Park*) bilden gewichtete Deskriptoren zur Verortung des Texts im Vektorraum. Im Kern ist dieses „*bag of words*“-Verfahren wortorientiert und strukturinsensitiv: Strukturkonstitutive Relationen bleiben ebenso ausgespart, wie die lexikalische Kohäsion von *Löwe*, *Skulptur* und *restaurieren*. Infolgedessen kommt es zu einer Verschmelzung der Vektorrepräsentationen dieser Wörter ohne Berücksichtigung der spezifischen Semantik ihrer lexikalischen Kohäsion.
2. Die LSA erweitert das Vektorraummodell, indem sie das wortbasierte Modell durch faktorenanalytische Repräsentationen ersetzt, die es erlauben, auch dann Ähnlichkeiten von Texten zu erkennen, wenn diese nicht überwiegend dieselben „wichtigen“, sondern ähnlich gebrauchte Wörter enthalten. Die „*bag of words*“-Problematik bleibt jedoch erhalten.
3. Mehler (2000) beschreibt eine Variante, bei der jede Komponente der Integrationshierarchie über die Menge ihrer lexikalischen Konstituenten auf den so genannten Textraum mittels eines Verfahrens abgebildet wird, das dieselben Ähnlichkeitsbeziehungen berücksichtigt, wie die LSA. Kohärenzrelationen und eine differenzierte Semantik der Integration bleiben jedoch ausgespart. Das Textbeispiel wird daher als ein Mengensystem von vektoriellen Wortrepräsentationen modelliert.

Das HCSP-Modell überwindet die Wortlastigkeit und Strukturindifferenz dieser Ansätze. Hier werden über Prädikat-Argument-Strukturen wie *restaurieren(Skulptur)* hinaus jene Kohärenzrelationen erfassbar, die neben der Integrationshierarchie obiges Textbeispiel konstituieren. Dies betrifft lexikalische Kohärenzrelationen ebenso wie die hypotaktische Relation von Nukleus und Satellit. Mehr noch, die Analyse von Prädikat-Argument-Strukturen bei Kintsch kann aufgrund der lexikalischen Kohärenz von

Löwen und *Skulpturen* in der Weise auf die Interpretation von *Löwen* übertragen werden, wie es dem Prinzip nach Kamp/Partee für das Kompositum *stone lion* vorsehen: *Löwen* wird (die Säugetierlesart unterdrückend) aus der Perspektive von *Skulpturen* interpretiert. Dies ist auch dann möglich, wenn im Beispiel die Satzreihenfolge umgedreht wird und *Löwen* über einen Revisionsprozess zu interpretieren ist.

Ganz analog ist die Interpretation des Satelliten durch jene Perspektive restringiert, welche die Interpretation des Nukleus stiftet: erstere Interpretation ist in Bezug auf die Interpretation des Gesamttextes letzterer gegenüber untergeordnet. Dabei ist zu berücksichtigen, dass *Park* je nach Repräsentation der Mehrdeutigkeit von Skulptur möglicherweise eine Revision (Rekalibrierung bei Kamp/Partee, 1995) der Interpretation von *Skulpturen* im Sinne von *Statuen*, *Plastiken* (und nicht im Sinne von *Bildhauerkunst*) bedingt. Freilich setzt dies (wie auch andere, hier nicht diskutierte Aspekte der Interpretation des Textbeispiels) die Spezifikation der Semantiken aller involvierten Kohärenzrelationen ebenso voraus wie ein hinreichend breites Korpus, anhand dessen die strukturellen Bedeutungen der Lexeme gelernt werden können. Betrachtet man die computerlinguistischen Fortschritte im Bereich von Anaphernresolution und Textparsing, so erscheint dies durchaus als realisierbar.

5 Schlussfolgerungen und Ausblick

In diesem Artikel wurde eine Variante des KPs für numerische Semantiken zur Repräsentation von Aspekten der strukturellen, konnotativen Bedeutung textueller Einheiten skizziert und mit Hilfe des Konzepts des HCSPs und des semiotischen Raums konkretisiert. Mit diesem Vorschlag ist der Versuch verbunden, im Bereich der verteilten, numerischen Repräsentation von gebrauchsbzw. inhaltsbasierten Ähnlichkeitsbeziehungen sprachlicher Einheiten zu einer Systematik zu gelangen, wie sie für die formale, logische Semantik kennzeichnend ist. Die Erörterungen dieses Artikels sind, da teils Anfänge betreffend, konzeptionell. Ihre empirische Evaluation auf der Basis des von Kintsch (2001) erprobten Verfahrens und ihre Modifikation im Lichte entsprechender Resultate bilden den nächsten Verfahrensschritt.

Die praktische Relevanz dieses Ansatzes ist eng mit der wachsenden Bedeutung von Verfahren zur automatischen Analyse von Zeichenrelationen verbunden. Dies kann am Beispiel der hypertextuellen Aufbereitung von Textkorpora motiviert werden. Anders als das *Information Retrieval* erweitert eine numerische Textsemantik, welche die Ähnlichkeitsbeziehungen von textuellen Einheiten genauer zu analysieren vermag, das Repertoire von Formaten zur kontextsensitiven Traversierung von Texträumen. Auch dies soll in nachfolgenden Untersuchungen thematisiert werden.

Literatur

- Barwise, J./Perry, J. (1983): *Situations and Attitudes*. Cambridge: MIT Press.
- Bierwisch, M. (1988): Auspizien einer kognitiven Orientierung der Computerlinguistik. In Bátori, I. S./Hahn, U./Pinkal, M./Wahlster, W. (Hrg.): *Computerlinguistik und ihre theoretischen Grundlagen*. Berlin [u.a.]: Springer, 198-203.
- Deerwester, S./Dumais, S. T./Furnas, G. W./Landauer, T. K./Harshmann, R. (1990): Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science* 41(6), 391-407.
- Dijk, T. A. van/Kintsch, W. (1983): *Strategies of Discourse Comprehension*. New York [u.a.]: Academic Press.
- Foltz, P. W. (1996). Latent Semantic Analysis for Text-Based Research. *Behavior Research Methods, Instruments & Computers* 28(2), 197-202.
- Halliday, M. A. K./Hasan, R. (1976): *Cohesion in English*. London: Longman.
- Janssen, T. M. V. (1997): Compositionality. In van Benthem, J./ter Meulen, A. (Hrg.): *Handbook of Logic and Language*. Amsterdam: Elsevier, 417-473.
- Kamp, H./Partee, B. (1995): Prototype Theory and Compositionality. *Cognition* 57(2), 129-191.
- Kintsch, W. (1998): *Comprehension. A Paradigm for Cognition*. Cambridge: University Press.
- (2001): Predication. *Cognitive Science* 25, 173-202.
- Knott, A./Sanders, T. (1998): The Classification of Coherence Relations and their Linguistic Markers: An exploration of two Languages. *Journal of Pragmatics* 30, 135-175.
- Lahav, R. (1989): Against Compositionality: the Case of Adjectives. *Philosophical Studies* 57(3), 261-279.
- Landauer, T. K./Dumais, S. T. (1997): A solution to Plato's problem. *Psychological Review* 104(2), 211-240.
- Mann, W. C./Thompson, S. A. (1988): Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8, 243–281.
- Martin, J. R. (1992): *English Text: System and Structure*. Philadelphia: John Benjamins.
- Mehler, A. (2000): *Textbedeutung. Zur prozeduralen Analyse und Repräsentation struktureller Ähnlichkeiten von Texten*. Frankfurt a. M.: Peter Lang. (Zugl. Diss. Universität Trier).
- (2002): *Textmining*. In Lobin, H./Lemnitzer, L. (Hrg.): *Texttechnologie. Perspektiven und Anwendungen*. Tübingen: Stauffenburg.
- Miller, G. A./Charles, W. G. (1991): Contextual Correlates of Semantic Similarity. *Language and Cognitive Processes* 6(1), 1-28.
- Osherson, D. N./Smith, E. E. (1981): On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition* 9(1), 35-58.
- Partee, B. H. (1984): Compositionality. In Landman, F./Veltman, F. (Hrg.): *Varieties of Formal Semantics. Proc. of the 4th Amsterdam Colloquium, Sept. 1982*. Dordrecht: Foris, 281-311.
- Renear, A./Mylonas, E./Durand, D. (1996): Refining our Notion of what Text really is: The Problem of Overlapping Hierarchies. In Ide, N./Hockey, S. (Hrg.): *Research in Humanities Computing*. Oxford: Oxford University Press, 263-280.

- Rieger, B. (1989): Unschärfe Semantik. Die empirische Analyse, quantitative Beschreibung, formale Repräsentation und prozedurale Modellierung vager Wortbedeutungen in Texten. Frankfurt a. M. [u.a.]: Peter Lang.
- Ruge, G. (1995): Wortbedeutung und Termassoziation. Methoden zur automatischen semantischen Klassifikation. Hildesheim [u.a]: Olms.
- Sanders, T./Spooren, W. (2001): Text Representation as an Interface between Language and its Users. In Sanders, T./Schilperoord, J./Spooren, W.: Text Representation: Linguistic and Psycholinguistic Aspects. Amsterdam: Benjamins, 1-25.
- Salton, G. (1989): Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Reading, Massachusetts [u.a.]: Addison Wesley.
- Schütze, H. (1998). Automatic Word Sense Discrimination. *Computational Linguistics* 24(1), 97-123.
- Sharkey, A. J. C./Sharkey, N. E. (1992): Weak Contextual Constraints in Text and Word Priming. *Journal of Memory and Language* 31(4), 543-572.
- Thagard, P. (2000): Coherence in Thought and Action. Cambridge: MIT Press.