

# Synonyme bei der Erzeugung, Generierung und Produktion natürlicher Sprache

Stephan Mehl

Gerhard-Mercator-Universität -GHS- Duisburg, Computerlinguistik,  
D-47048 Duisburg

**Zusammenfassung.** Für eine abwechslungsreiche Wortwahl ist die Verwendung von Synonymen unerlässlich. Allerdings sind Synonymierelationen wegen ihrer Kontextabhängigkeit in traditionellen Lexika nur schwer exakt darstellbar. Außerdem tragen nicht alle Synonyme in gleicher Weise zur Lesbarkeit und Verständlichkeit eines Textes bei. Zur Berücksichtigung der Kontextabhängigkeit wird ein Verfahren vorgestellt, das anhand von Aktivationsausbreitungsprozessen auf einem semantischen Netz verschiedene Formulierungsalternativen vorschlägt. Sodann werden Kriterien diskutiert, die eine Auswahl unter diesen Alternativen ermöglichen.

**Abstract.** Varying words in natural language generation requires the use of synonyms. In traditional dictionaries, however, the formal representation of synonyms is difficult due to their dependence on context. In addition to that, not all synonyms enhance the readability and comprehensibility of a text equally well. It is proposed to take the context of synonyms into account by using a spreading activation mechanism in a semantic network. Several criteria are discussed to choose among the so-obtained alternatives.

## 1. Problemstellung

Was die Rolle des Lexikons betrifft, so besteht das Hauptproblem bei der maschinellen Generierung natürlicher Sprache seit jeher darin, für einen vorgegebenen Inhalt das passende Wort zu finden. Dies gilt für alle Bereiche der Sprachgenerierung, von Informationssystemen bis hin zur Maschinellen Übersetzung. In diesem Beitrag sollen die Anforderungen an ein gutes System noch etwas höher geschraubt werden: Wäre es nicht wünschenswert, statt nur eines einzigen inhaltlich passenden Wortes mehrere zur Verfügung zu haben, aus denen man das syntaktisch oder stilistisch geeignetste auswählen kann? Bei der Generierung von Antworten in einem Dialogsystem könnte dann z.B. statt

*Das gesuchte Buch befindet sich im Handapparat Informatik. Dieser befindet sich in Raum LF245. Das Gebäude LF befindet sich an der Lotharstraße,*

zwischen der Bibliothek und der Mensa. Ein Informationsblatt mit den genauen Benutzungsbedingungen befindet sich an der Tür des Raums LF 244.

eher der folgende Text erzeugt werden:

*Das gesuchte Buch steht im Handapparat Informatik. Dieser ist in Raum LF245 untergebracht. Das Gebäude LF liegt an der Lotharstraße, zwischen der Bibliothek und der Mensa. Ein Informationsblatt mit den genauen Benutzungsbedingungen hängt an der Tür des Raums LF 244 aus.*

Im Bereich der Verben ist die Notwendigkeit besonders groß, Synonyme zur Verfügung zu haben, weil hier in weitaus geringerem Maße als bei den Nomina Pro-formen verwendet werden können. Allerdings ist es nicht damit getan, einfache Listen von synonymen Verben ins Lexikon aufzunehmen. Einige der mit der Ermittlung von Synonymen verbundenen Schwierigkeiten werden im folgenden in Abschnitt 2 diskutiert.

Neben der stilistischen Variation gibt es übrigens auch andere Gründe dafür, daß ein Synonym besser als ein anderes geeignet sein könnte, einen bestimmten Inhalt auszudrücken. So können sich verschiedene Verben darin unterscheiden, wie sie dieselben Kasusrollen auf Valenzstellen abbilden. Wie in Abschnitt 3 gezeigt wird, können dieser und andere Unterschiede dazu führen, daß das eine Verb zu mißverständlicheren oder schwerer lesbaren Texten führt als das andere. - Abschnitt 4 schließlich zeigt Möglichkeiten auf, die in den Abschnitten 2 und 3 beschriebenen Desiderate in einer Implementierung zu realisieren.

## 2. Ermittlung von Synonymen

Die genaue Repräsentation von lexikalischen Informationen in einem Sprachgenerierungssystem hängt von dessen Aufgabenstellung und Architektur ab. Grob vereinfachend, enthalten alle Systeme eine Transferkomponente, die die Ausdrücke einer natürlichen Quellsprache oder einer formalen semantischen Repräsentations-sprache auf Lexeme der Zielsprache abbildet. An dieser Stelle wäre die Existenz alternativer Synonyme festzustellen. Um von der konkreten Gestalt dieser Transferkomponente zu abstrahieren, wird im folgenden einfach davon ausgegangen, es seien Synonymiebeziehungen in einem umfassenden Lexikon (des Deutschen) dar-zustellen, unabhängig von deren Beziehung zu einer inhaltlichen Vorgabe.

Das zentrale Problem bei der Darstellung von Synonymierelationen besteht in der semantischen Kontextabhängigkeit der Synonymie. Im obigen Beispiel wurde *sich befinden* ersetzt durch *stehen*, *untergebracht sein*, *liegen* und *aushängen*. Diese Lexeme sind jedoch untereinander keineswegs synonym, sie stellen nur in speziellen Kontexten eine Alternative zu *sich befinden* dar: ein Gebäude, das sich an einem bestimmten Ort *befindet*, *liegt* dort, ein Buch *steht*, ein Informationsblatt an einer Tür *hängt aus*, ein Informationsblatt auf einem Tisch *liegt aus* etc. Bei einer engen Auslegung des Polysemiebegriffs könnte man sagen, daß sich die Bedeutung von *sich befinden* je nach Kontext leicht verschiebt, was sich an der unterschiedlichen Paraphrasierbarkeit durch Synonyme zeigt (vgl. Wichter 1988). Alternativ könnte man eine "Kernbedeutung" von *sich befinden* ("an einem bestimmten Ort sein") annehmen und argumentieren, daß es etwa von außersprachlichem Wissen

über die Position des Informationsblatts abhängt, welches Verb (*aushängen* vs. *ausliegen*) als Paraphrase benutzt werden kann. In beiden Fällen kann das Lexikon nicht einfach eine Reihe von Synonymen auflisten, sondern es sind Kontextangaben nötig, um die jeweils passenden ermitteln zu können. Im einfachsten Fall handelt es sich dabei um Kollokationsangaben, die explizit in Lexika verzeichnet sind, weil bei ihnen das abhängige Element eine spezielle, nur auf das dominierende Element bezogene Bedeutung bekommt (z.B. *die Uhr steht*, vgl. Hausmann 1985).

Im folgenden wird der Begriff der Synonymie so weit gefaßt, daß auch solche Verben darunter fallen, die zwar dasselbe Prädikat über denselben Argumenten ausdrücken, den semantischen Kasusrahmen syntaktisch jedoch völlig unterschiedlich realisieren. Beispiel: *wir haben nichts mit dieser Angelegenheit zu tun* - *diese Angelegenheit betrifft uns nicht*.

Das Paar *betreffen* - *mit etwas zu tun haben* ist ein Beispiel dafür, daß häufig nicht zwei Lexeme derselben Wortart synonym sind, sondern zwei Syntagmen, die möglicherweise eine unterschiedliche syntaktische Gestalt haben. Hierzu gehören feste Syntagmen (*planen* - *im Schilde führen*) ebenso wie Funktionsverbgefüge: *sich auswirken* könnte durch *Auswirkungen haben* und dieses wiederum durch *Folgen haben* ersetzt werden (dagegen ist *sich auswirken* nicht synonym zu *folgen*). Um eine größere Freiheit bei der Textplanung zu erreichen, ist es sinnvoll, auch Nominalisierungen mit einzubeziehen: Inhaltlich gesehen, besteht kein Unterschied zwischen *die Auswirkungen dieser Veränderung auf die Distribution* und *diese Veränderung wirkt sich auf die Distribution aus*. Stilistisch sind Nominalisierungen zwar verpönt, sie sind aber nützlich zur Erreichung einer knapperen Darstellung eines Sachverhalts durch Vermeidung der Aneinanderreihung von Nebensätzen.

Über die Verwendung von Synonymen hinaus gibt es weitere Mittel, um Abwechslung in die Bezeichnung desselben Gegenstandes bzw. derselben Handlung zu bringen. Ein gängiges Mittel ist die Verwendung von Ober- oder von Unterbegriffen; bei den Unterbegriffen muß natürlich kontrolliert werden, ob sie zutreffen, und die Oberbegriffe könnten zu Verwechslungen mit anderen möglichen Referenzobjekten führen. Im Rahmen dieser Arbeit wird daher nur der einfachere Fall der Synonymie betrachtet.

### 3. Auswahl aus verschiedenen Alternativen

Angenommen, für einen gegebenen Kontext seien mehrere alternative Synonyme ermittelt worden; z.B.

*das Buch steht im Handapparat/ist Teil des Handapparats/ist im Handapparat zu finden*

*das Gebäude liegt an der Lotharstraße/ist an der Lotharstraße gelegen;*

*das Informationsblatt hängt aus/ist ausgehängt/ist angeschlagen*

Im folgenden werden Kriterien vorgeschlagen, nach denen eines von mehreren für einen bestimmten Kontext ermittelten Synonymen ausgewählt werden kann.

#### 3.1 Allgemeine Kriterien

Stilistische Angemessenheit: Synonyme, die einer unpassenden Stilebene, einem bestimmten Soziolekt etc. angehören, sollten ausgeschlossen werden. Oft handelt es sich hier allerdings nicht um binäre, sondern um graduelle Bewertungen: Verglichen mit der Formulierung

*ein Informationsblatt mit den Benutzungsbedingungen hängt an der Tür aus*  
wäre *ein Blatt mit den Benutzungsbedingungen hängt an der Tür salopper*, die *Benutzungsbedingungen werden durch Aushang an der Tür bekanntgegeben* formeller. Wenn aber nicht gerade eine saloppe und eine formelle Formulierung miteinander kombiniert werden, wären beide Varianten in dem obengenannten Text durchaus noch akzeptabel.

Syntaktische Komplexität: Durch ihre Gestalt ergeben sich bei manchen Lexemen syntaktisch komplexere Strukturen als bei anderen. So ist etwa die Aktivform *das Gebäude liegt* dem Zustandspassiv *das Gebäude ist gelegen* vorzuziehen (wobei in diesem Beispiel noch stilistische Unterschiede hinzukommen). *Das Buch ist im Handapparat des Instituts für Informatik im Gebäude LF an der Lotharstraße zu finden* wird durch die späte Nennung des Verbs schwerer verständlich als *Das Buch steht im Handapparat des Instituts für Informatik im Gebäude LF an der Lotharstraße*.

Syntaktisch unterschiedliche (und damit eventuell unterschiedlich leicht verständliche) Strukturen ergeben sich auch aus der unterschiedlichen Zuordnung der semantischen Kasus- zu den syntaktischen Valenzrahmen bei verschiedenen Lexemen. So unterscheiden sich *aushängen* und *anschlagen* dadurch, daß beide zwar einen Valenzrahmen aufweisen, bei dem das Patiens durch eine Akkusativergänzung wiedergegeben wird, *aushängen* darüber hinaus in derselben Bedeutung aber noch einen zweiten Valenzrahmen mit einem Patiens im Nominativ aufweist. Es erlaubt daher auch dann eine Aktivkonstruktion, wenn kein Agens zur Verfügung steht:

*Das Informationsblatt hängt an der Tür aus.*

*Das Informationsblatt ist an der Tür ausgehängt.*

*\*Das Informationsblatt schlägt an der Tür an.* (vgl. *die Tür schlägt an der Wand an*)

*Das Informationsblatt ist an der Tür angeschlagen.*

Das heißt nicht, daß eine Passivformulierung automatisch abzulehnen wäre; aus Variationsgründen kann sie durchaus angebracht sein, wenn nicht weitere Faktoren gegen sie sprechen (z.B. ein umfangreiches Mittelfeld wie oben im *Handapparat*-Beispiel, das Subjekt und Verb zu weit voneinander trennt). Letzteres kann jedoch erst im Zuge der syntaktischen Generierung entschieden werden.

### 3.2 Präferierte Valenzrahmen

In verschiedenen sprachpsychologischen Untersuchungen (beginnend mit Ford et al. 1982 sowie Connine et al. 1984) wurde festgestellt, daß nicht alle Valenzrahmen, die ein bestimmtes Verb zuläßt, denselben Stellenwert haben. So kann ein bestimmtes Komplement beispielsweise optional sein, aber dennoch mehr oder weniger stark erwartet werden. Anders ausgedrückt: Ein Verb mit einer optionalen Präpositionalergänzung läßt zwei verschiedene Valenzrahmen zu, einen mit und einen

ohne diese Präpositionalergänzung; diese beiden Valenzrahmen können unterschiedlich hohe Wahrscheinlichkeitswerte haben. Diese Wahrscheinlichkeitswerte äußern sich in der Interpretation bestimmter syntaktisch mehrdeutiger Sätze. Beispiel (Ford et al. 1982): Sowohl in

*The woman wanted the dress on that rack.*

als auch in *The woman positioned the dress on that rack.*

kann *in the rack* als Ergänzung des Verbs oder als Attribut zu *dress* interpretiert werden. Einer Gruppe von Versuchspersonen wurden diese und ähnliche Sätze zusammen mit zwei möglichen Paraphrasen vorgelegt mit der Bitte, anzugeben, welche der beiden Interpretationen ihnen zuerst eingefallen war. Bei *want* entschieden sich 90% der Versuchspersonen für die Attributlesart (*The woman wanted the dress which was on that rack*) und 10% für die Komplementlesart (*The woman wanted it (the dress) on that rack*). Bei *position* war das Verhältnis umgekehrt (30% für die Attribut-, 70% für die Komplementlesart). - Ähnliche Präferenzen wurden nicht nur bei Valenzrahmen, sondern auch bei Kasusrahmen nachgewiesen (Pritchett 1987).

Natürlich können solche lexikalischen Präferenzen durch andere Faktoren annulliert werden (etwa wenn der erste Satz eine Antwort auf die Frage *Where shall I put the dress?* darstellt). Falls solche Faktoren jedoch fehlen oder erst später nachgeschoben werden, führen sie zur Konstitution einer bestimmten Interpretation, die möglicherweise später mühsam revidiert werden muß. Zur Vermeidung von falschen Interpretationen sollten solche Interpretationsstrategien daher bei der Generierung antizipiert werden.

Bei *want* wird die Situation dadurch erschwert, daß die beiden Valenzrahmen zwei verschiedenen Lesarten zuzurechnen sind: In der Attributlesart bedeutet *want* "besitzen wollen", in der Komplementlesart "wollen, daß etwas geschieht" (*The woman wanted the dress to be positioned on the rack.*) Von daher fällt *want* eher unter die nun folgende Diskussion der Lesartenpräferenz bei Polysemen.

### 3.3 Mehrdeutige Wörter

Wesentlich deutlicher als die in 3.2 beschriebenen lexikalischen Präferenzen sind die Unterschiede, die zwischen den einzelnen Lesarten eines Homonyms oder Polysems bestehen. Die Begründung für die Relevanz dieser Unterschiede ist dieselbe wie in 3.2: Sind in der Menge der Synonyme, von denen eines auszuwählen ist, solche, die in einer unpräferierten Lesart verwendet werden müßten, so sind diejenigen vorzuziehen, bei denen die präferierte Lesart zum Tragen kommt. Auf diese Weise lassen sich Mißverständnisse nach Möglichkeit vermeiden.

Eine bestimmte Lesart kann aus verschiedenen Gründen präferiert werden (vgl. Small/Cottrell/Tanenhaus 1988): weil sie grundsätzlich am häufigsten benutzt wird ("dominance"), weil sie am besten zum jeweiligen Kontext paßt ("salience") oder weil das betreffende Wort kürzlich schon in dieser Lesart verwendet wurde ("recency"). Nach Möglichkeit sollten Homonyme und Polyseme nur dann bei der Generierung eingesetzt werden, wenn die verwendete Lesart allen drei Kriterien genügt oder wenn das zutreffende Kriterium so stark ist, daß die übrigen keine Rolle mehr spielen. Einige Beispiele:

Dominanz: Vom Informationsgehalt her dürften die beiden folgenden Formulierungen annähernd gleich sein:

*Ein Anschlag neben der Bibliothek informiert über eventuelle Änderungen. Informationen über eventuelle Änderungen sind neben der Bibliothek angeschlagen.*

Die beiden Sätze unterscheiden sich jedoch dadurch, daß bei dem oberen gleich zwei Lexeme in einer ungewöhnlichen Lesart verwendet werden, während bei dem unteren lediglich die dominanten Lesarten zum Tragen kommen. *Anschlag* wird meist in der Bedeutung *Angriff benutzt*, und bei *informieren* wird das Subjekt üblicherweise eher durch das Agens denn durch das Instrument gebildet. Die geringe Dominanz der hier intendierten Lesart von *Anschlag* wird zwar durch deren große Salienz kompensiert; trotzdem wirkt der zweite Satz verständlicher als der erste.

Salienz: Der *Handapparat*, in dem *das gesuchte Buch steht*, ist selbstverständlich eine Handbibliothek und nicht etwa ein Telefonhörer (oder was sonst noch in einem technischen Fachwörterbuch als Bedeutung dieses Wortes verzeichnet sein mag). Je nach Kontext sind bestimmte Bedeutungen eines Wortes eher zu erwarten als andere, unabhängig von den klassischen Selektionsrestriktionen, nach denen ein Verb möglicherweise nur mit bestimmten Lesarten eines polysemen Nomens kombinierbar ist.

Recency: Wird ein Polysem zweimal kurz nacheinander verwendet, so ist zu erwarten, daß es in beiden Fällen dieselbe Bedeutung hat. Texte wie der folgende sollten daher vermieden werden:

*Das Buch befindet sich im Handapparat Informatik. Dieser befindet sich jedoch derzeit in Auflösung, so daß das Buch erst in der nächsten Woche zur Verfügung stehen wird.*

#### 4. Implementierung

In Abschnitt 3 wurden verschiedene Kriterien dargestellt, nach denen aus einer Liste von Synonymen eine Auswahl getroffen werden kann. Viele dieser Kriterien liefern graduelle statt binärer Bewertungen, und viele stehen in wechselseitigen Abhängigkeitsverhältnissen. Abschnitt 2 hob das Problem hervor, angesichts der Kontextabhängigkeit von Wortbedeutungen überhaupt erst zu Listen von Synonymen zu kommen. Im folgenden wird eine Implementierung dargestellt, die beide Problemfelder zu erfassen versucht.

Das Problem der Kontextabhängigkeit der Wortbedeutungen wird ausführlich in Mehl 1993 diskutiert. Dort wird vorgeschlagen, das Bedeutungsspektrum eines Polysems nicht künstlich in verschiedene Lesarten aufzuteilen, sondern durch ein Netz von semantischen Relationen darzustellen. Nach diesem Modell soll die kontextabhängige Konstitution einer Wortbedeutung durch einen Aktivationsausbreitungsprozeß auf diesem Netz repräsentiert werden, bei dem die von einem Wortknoten ausgehenden semantischen Relationen ein unterschiedlich starkes Gewicht erhalten.

Für unsere Zwecke sind besonders die Synonymierelationen interessant, die Bestandteil dieses Netzes sind. Abb. 1 stellt einige dieser Synonymien in stark vereinfachter Form dar, zusammen mit syntagmatischen Kasusrelationen und Hyponymien. Aus Gründen der Übersichtlichkeit sind andere Kanten desselben Netzes in Abb. 2 getrennt aufgeführt, nämlich solche, in denen *sitzen*, *stehen* und *liegen* nicht synonym zu *sich befinden* sind, sondern hyponym ("sich in einer bestimmten Stel-

lung befinden"). So kann man von einem Buch auch dann sagen, es stehe in der Bibliothek (sei Bestandteil dieser Büchersammlung), wenn es tatsächlich in dem Raum, in dem diese Büchersammlung untergebracht ist (der sog. Bibliothek), auf dem Tisch liegt. Das in Mehl 1993 beschriebene Netz repräsentiert insbesondere die Bedeutungsverwandtschaft zwischen diesen Verwendungsweisen von *stehen* etc., die sich z.B. darin ausdrückt, daß eine Äußerung wie

*Dieses Buch steht im Handapparat, und zwar im linken Regal oben.*  
als wohlgeformt und nicht als zeugmatisch betrachtet wird.

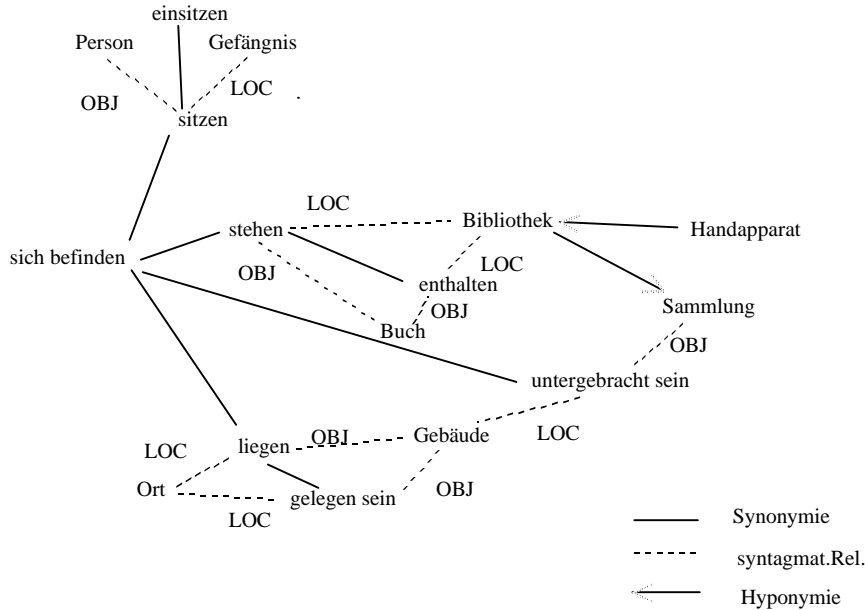


Abb. 1

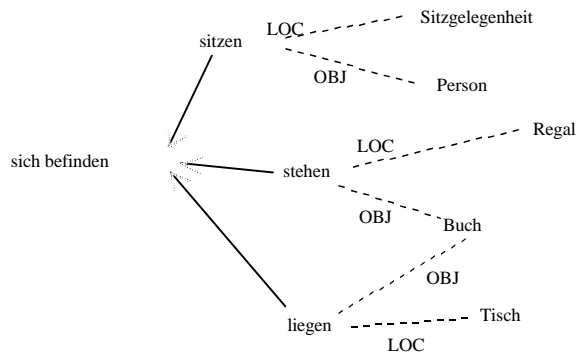


Abb. 2

Der Aktivationsausbreitungsprozeß bewirkt, daß diejenigen semantischen Relationen, die auf Pfaden zwischen den Eingabewörtern liegen, gegenüber den anderen Relationen hervorgehoben werden. Unterschiedliche Kombinationen von Eingabewörtern führen damit zu unterschiedlichen Aktivationsmustern. Eine Synonymiere-lation wird nur dann als in einem bestimmten Kontext vorliegend angesehen, wenn sie durch die Eingabe der Wörter dieses Kontexts aktiviert wird. Ob z.B. die Formulierung *ein Buch steht* als synonym oder als hyponym zu *ein Buch befindet sich* betrachtet werden muß, hängt davon ab, ob als Ortsangabe *Bibliothek* oder *Regal* etc. genannt wird. Der Aktivationsausbreitungsmechanismus sorgt dafür, daß die Synonymie- bzw. Hyponymiekanten zwischen *stehen* und *sich befinden* entsprechend mehr oder weniger stark aktiviert werden. Umgekehrt trägt *Bibliothek* nur in Verbindung mit *Buch* zur Aktivierung der Synonymiekante bei; in Verbindung etwa mit *Staubsauger* (*Der Staubsauger steht in der Bibliothek.*) wird *stehen* als hyponym zu *sich befinden* erkannt. - Ein zweites Beispiel: *untergebracht sein* zählt nur in Verbindung mit *Sammlung* o.ä. als synonym zu *sich befinden*, nicht aber z.B. in Verbindung mit *Hotelgast* (ein in einem Hotelzimmer untergebrachter Hotelgast befindet sich trotzdem nicht ständig dort).

Die unterschiedlich starke Aktivierung einer Relation durch den Kontext reflektiert die Salienz der durch diese Relation repräsentierten Lesart. Das Aktivationsniveau einer Kante bleibt über einen gewissen Zeitraum erhalten, bevor es langsam absinkt; damit wird auch dem Recency-Effekt Rechnung getragen. Die Dominanz einer Lesart schließlich wird durch eine spezielle Gewichtung der Kanten dargestellt, die in den Aktivationsausbreitungsprozeß eingeht. So erhält beispielsweise die Kante *sitzen - Gefängnis* eine deutlich niedrigere Bewertung als *sitzen - Sitzgelegenheit*; dies bedeutet, daß es einer ausdrücklichen Erwähnung von *Gefängnis* (oder zumindest von *Straftat*, *Verbrecher* o.ä.) bedarf, um das Verb *sitzen* als "im Gefängnis sitzen" zu interpretieren. Ohne eine solche explizite Kontextangabe wird *sitzen* als "auf einer Sitzgelegenheit sitzen" verstanden (z.B. in *Der moderne Mensch sitzt zuviel.*).

Die exakten numerischen Bewertungen der Kanten nach Ablauf der Aktivationsausbreitung hängen von einer Reihe von Parametern wie z.B. der Netzgröße ab (vgl. Mehl 1993). Im folgenden werden diese Bewertungen daher zu Vergleichszwecken auf eine Skala von 0 bis 10 abgebildet.

Um die übrigen in Abschnitt 3 genannten Kriterien zur Auswahl zwischen verschiedenen Synonymen einsetzen zu können, werden die Knoten folgendermaßen annotiert:

- stilistische Neutralität: eine Bewertung zwischen 0 (= zu salopp oder zu formell) und 3 (=neutral)
- Allgemeinverständlichkeit: eine Bewertung zwischen 0 (nur in Fachtexten verwendbar) und 3 (=allgemein bekannt). Bei Bewertungen <3 wird zusätzlich vermerkt, zu welcher Fachsprache das Wort gehört.
- Valenzrahmen: Jeder Verbknoten erhält eine Liste der jeweils möglichen Valenzrahmen, versehen mit einer Präferenzbewertung zwischen 0 (= möglichst zu vermeiden) und 3 (= nach Möglichkeit zu verwenden). Für jeden Valenzrahmen werden die zugehörigen möglichen Kasusrahmen angegeben, ebenfalls jeweils mit einer Präferenzbewertung versehen. Aufgrund der Zuordnung von Kasus- zu Valenzrahmen ist es beispielsweise möglich, die Kanten (*enthalten OBJ: Buch*,



LOC: *Bibliothek*) durch *Die Bibliothek enthält das Buch*, die synonyme Formulierung (*stehen* OBJ: *Buch*, LOC: *Bibliothek*) durch *Das Buch steht in der Bibliothek* zu realisieren.

Schließlich geht in die Entscheidung noch die Komplexität der erzeugten syntaktischen Struktur ein (z.B. Vermeidung von Passivformulierungen). Zu Demonstrationszwecken wird diese Komplexität im folgenden Beispiel mit einer Bewertung zwischen 0 und 6 versehen.

Auf dieser Basis läuft die Generierung verschiedener Formulierungen folgendermaßen ab:

Im Zuge der Inhaltsfestlegung werden Konzepte bestimmt, die auf ausgewählte Lexeme und deren Rollen abgebildet werden. Beispiel: (*sich befinden* OBJ: *Buch*, LOC: *Handapparat*), Fokus auf *Buch*. Ausgehend von den drei Knoten, die diesen Lexemen zugeordnet sind, wird ein Aktivationsausbreitungsprozeß in Gang gesetzt, der als Alternative zu *sich befinden* die Synonyme *stehen* und *enthalten* ermittelt. In diesem einfachen Beispiel kommt bei jedem der Lexeme nur ein Valenzrahmen in Betracht. Die Einzelbewertungen ergeben folgendes Resultat:

*sich befinden*: stilistische Neutralität 2, Allgemeinverständlichkeit 3, Valenz- und Kasusrahmen 3. Verb und Komplemente passen gut zusammen, *Das Buch befindet sich* läßt jedoch mehrere Interpretationen offen (*das Buch befindet sich auf dem Weg zum Buchbinder/in schlechtem Zustand* etc.), von denen keine dominant ist. Kantenbewertung daher: 8. Syntaktische Realisierung mit dem erwünschten Fokus: 5 (Punktabzug wegen Reflexivität). Gesamtpunktzahl: 21.

*stehen*: stilistische Neutralität 3, Allgemeinverständlichkeit 3, Valenz- und Kasusrahmen 3. Verb und Komplemente passen gut zusammen, *Das Buch steht* läßt eine weitere Interpretationen offen (*das Buch steht im Regal*), die aber eng mit der hier intendierten verwandt ist. Kantenbewertung 10. Syntaktische Realisierung mit dem erwünschten Fokus: 6. Gesamtpunktzahl: 25.

*enthalten sein*: stilistische Neutralität 2, Allgemeinverständlichkeit 3, Valenz- und Kasusrahmen 3. Verb und Komplemente passen gut zusammen, keine weiteren Interpretationsmöglichkeiten. Kantenbewertung 10. Syntaktische Realisierung mit dem erwünschten Fokus: 3 (wegen des Fokus Passivformulierung: *das Buch ist im Handapparat enthalten*). Gesamtpunktzahl: 21.

Gesamtbewertung: Als präferierte Erstreferenz ergibt sich *Das Buch steht im Handapparat*. Als Folgereferenz kann eine der beiden anderen Formulierungen verwendet werden (z.B. *es befindet sich dort noch bis zum Ende des Sommersemesters; der Handapparat befindet sich ...*).

## 5. Ausblick

Die in Abschnitt 4 verwendeten numerischen Bewertungen sollen lediglich einen Eindruck davon geben, wie die Entscheidung für oder gegen eine bestimmte Formulierung vom Wechselspiel verschiedener Einflußgrößen abhängt. Zur Zeit laufen genauere Untersuchungen zur Modellierung dieses Wechselspiels. Aus dem in

Mehl 1993 dokumentierten semantischen Netz stehen hierzu nur ca. 130 Synonymierelationen zur Verfügung. Dieses Inventar wird jedoch ergänzt durch ca. 6000 Synonymien, die automatisch aus dem Duden - Deutschen Universalwörterbuch extrahiert wurden (Mehl 1994) und an ausgewählten Stellen gezielt manuell ergänzt werden, vor allem durch komplexere Syntagmen.

Leider beschäftigen sich nur wenige Veröffentlichungen mit der Problematik der parallelen Untersuchung von Formulierungsalternativen. Besonders im lexikalischen Bereich geben sich viele Verfahren mit der erstbesten Lösung zufrieden - wenn überhaupt Alternativen vorgesehen sind. Für den Bereich der Syntax beschreibt Ward 1992 ein Verfahren, das die Auswahl einer syntaktischen Konstruktion ebenfalls mit Hilfe eines Aktivationsausbreitungsmechanismus steuert. Wie oben in Abschnitt 3 dargestellt wurde, ist die Interaktion von lexikalischen und syntaktischen Realisierungsentscheidungen ein zentrales, wenn auch schwer zu realisierendes Element des Generierungsprozesses. - Hovy 1988 untersucht ebenfalls Formulierungsalternativen unter Verwendung einer Architektur ineinander verzahnter Prozesse, konzentriert sich aber auf solche Varianten, die unterschiedliche pragmatische Vorgaben realisieren (z.B. verschiedene Wertungen ausdrücken, emotional unterschiedlich gefärbt sind etc.).

Um eine weitergehende Vielfalt von Formulierungen zu ermöglichen, muß das Lexikon in eine Repräsentation von Handlungszusammenhängen eingebettet werden, die über das hier beschriebene semantische Netz weit hinaus geht. Hier wird das Problem angesprochen, wie Textplanung und Realisierung, denen die stilistische Ausgestaltung obliegt, und die Inhaltsfestlegung miteinander zu koordinieren sind.

## Literatur:

1. Connine, C./Ferreira,F./Jones,C./Clifton, C./Frazier,L. 1984: Verbframe preferences: Descriptive norms. *Journal of Psycholinguistic Research* 13, S. 307-319.
2. Ford, Marilyn/Bresnan,Joan/Kaplan,Ronald 1982: A Competence-Based Theory of Syntactic Closure. In: J. Bresnan (Hrsg.): *The mental representation of grammatical relations*. Cambridge, MA: MIT Press, S. 727-796.
3. Hausmann, Franz Josef 1985: Kollokationen im deutschen Wörterbuch. In: H. Bergenholz/J.Mugdán (Hrsg.): *Lexikographie und Grammatik*. Tübingen: Niemeyer, S. 118-129.
4. Hovy, Eduard H. 1988: *Generating natural language under pragmatic constraints*. Hillsdale: Lawrence Erlbaum.
5. Mehl, Stephan 1993: *Dynamische semantische Netze. Zur Kontextabhängigkeit von Wortbedeutungen*. St. Augustin: infix.
6. Mehl, Stephan 1994: Automatic acquisition of lexical-semantic relations from machine-readable dictionaries. In: René Dirven/Johan Vanparys (Hrsg.): *New Approaches to the Lexicon*. Frankfurt: Peter Lang.
7. Pritchett, Bradley L. 1987: *Garden path phenomena and the grammatical basis of language processing*. Diss. Harcard, Univ., University Microfilms.
8. Small, Steven/Cottrell, Garrison/Tanenhaus,Michael (Hrsg.) 1987: *Lexical Ambiguity Resolution*. Hillsdale: Lawrence Erlbaum.
9. Ward, Nigel 1992: A parallel approach to syntax for generation. *Artificial Intelligence* 57, S. 183-225.

10. Wichter, Sigurd 1988: Signifikantgleiche Zeichen. Tübingen: Narr.