

Referenzstrukturen. Zur Quantifizierung der Koinzidenz semantischer Einheiten in Texten

Verfolgt man die Idee, dass Sprachwissenschaft mehr zu leisten hat, als nach abstrakten Eigenschaften sprachlicher Ausdruckseinheiten zu fragen und diese zu beschreiben, gelangt man zwangsläufig in die turbulenten Bereiche semantischer Sprachbetrachtung. Ist dies der Fall, sieht man sich zwei fundamentalen Schwierigkeiten gegenüber: Zum einen verkompliziert sich die Sache, wenn festgestellt wird, dass Bedeutung nicht in den Wörtern oder Sätzen (allein) steckt, sondern sich aus Sprachwissen, Voreinstellungen, Intentionen, Kontextbedingungen, kommunikativen Handlungszusammenhängen usw. im jeweiligen konkreten Text konstituiert. Zum anderen sind aber eben diese Faktoren einer pragmatisch und semantisch orientierten Linguistik dem Wissenschaftler nicht unmittelbar zugänglich, sondern können nur indirekt über den Text ermittelt werden. Eine zentrale Frage einer eingangs postulierten Sprachbetrachtung ist daher: Mit welchen Methoden – die über den Rahmen einer rein deskriptiven Sprachbetrachtung hinausgehen – können wir uns einem Text nähern, um Kenntnisse hinsichtlich seiner semantischen Struktur zu erhalten?

Basierend auf der Methode der Denotativen Textanalyse¹ – wo diskurssemantische Einheiten – sogenannte *Hrebs* – als sprachliche Einheiten etabliert wurden – soll im Folgenden zunächst exemplarisch gezeigt werden, wie Referenzbeziehungen in Texten auf denotativer Grundlage ermittelt werden können. In einem zweiten Schritt wird vorgeführt, wie sich auf der Grundlage der denotativen Einheiten schließlich eine assoziative Struktur von Texten ermitteln lässt, die allen Ansprüchen an Operationalisierbarkeit und Validität genügt und damit auch für umfangreiche korpusbasierte Untersuchungen sowie für kontrastive Analysen geeignet scheint.

Es wird ein Modell der linguistischen Textstrukturanalyse dargestellt – in Anbetracht des zur Verfügung stehenden Rahmens natürlich nur sehr ausschnitthaft – dessen Erkenntnisinteresse nicht mehr ontologisch, sondern vorwiegend methodologisch orientiert ist und das versucht, Ordnungsmuster in Texten zu ermitteln, von denen anzunehmen ist, dass sie sich in einem kognitiven Differenzierungswissen der Sprachbenutzer manifestieren. Begriffe stellen nicht die Realität dar, sondern unsere Konstruktion der Realität.²

¹ Vgl. Ziegler 2001, 2003; Ziegler/Altmann 2002;

² Vgl. Rapoport 1972; Korzybski 1973; Köller 1988; Schmidt 1994; Ziegler 2001

1 Denotation und Referenz

Die Versuche der Linguistik – und insbesondere der Textlinguistik – semantische Einheiten in Texten zu etablieren sind vielfältig.³ Ausgehend von der Annahme, dass ein jeglicher Text aus kohärenten Elementen besteht, bleibt die Frage, worin diese Kohärenz besteht und welche kohäsiven Mittel dominant verantwortlich sind, auch gegenwärtig noch eine der viel diskutierten in der textlinguistischen Forschung.⁴ Weitgehend Einigkeit herrscht dagegen hinsichtlich der Auffassung, dass sowohl zwischen den Entitäten eines Textes untereinander sowie zwischen den Entitäten eines Textes und außersprachlichen Faktoren zahlreiche Relationen bestehen. Die semantischen Relationen stellen dabei nur eine Möglichkeit dar (andere Möglichkeiten sind phonetische Relationen, morphologische, grammatische usw.). Im Lexikon lassen sich semantische Relationen durch Begriffe wie *Synonymie*, *Polysemie*, *Hyponymie* usw. erfassen. Im konkreten Text haben diese Relationen aber eine ganz andere Beschaffenheit. Hier gruppieren sich Entitäten, die aufeinander referieren. Die Relationen der semantischen Entitäten können somit als Referenzbeziehungen verstanden werden. In diesem Sinne stehen Text und Referenz in unmittelbarem Zusammenhang: Im Text wird referiert und Referieren geschieht üblicherweise in Texten. Die Gesamtheit der semantischen Referenzbeziehungen eines Textes wird im Allgemeinen als *Isotopie* bezeichnet. Die Isotopie (oder auch die Isotopiekette) begründet somit die semantische Kohärenz eines Textes.⁵

Aus der Vielfalt aller möglichen textinternen semantischen Relationen werden in dem hier dargestellten Ansatz zunächst ausschließlich die denotativen Relationen fokussiert, d.h. aus dem Bereich der möglichen Isotopie eines Textes findet eine Beschränkung auf die denotativen Referenzbeziehungen statt, nicht zuletzt, um auf diese Weise eine möglichst zweifelsfreie Operationalisierbarkeit zu gewährleisten. In diesem Zusammenhang ist die Frage nach Möglichkeiten und Formen der denotativen Kohärenzstiftung eines Textes von zentraler Bedeutung.⁶

Text ist ein in der Zeit verlaufender Prozess, in dem es zwei Ströme gibt. Einen geordneten, laminaren Strom der Formen, der – besonders in geschriebenen Texten –

³ Vgl. z.B. Agricola 1969, Viehweger 1976, 1977, 1978; van Dijk 1977; Lerchner 1984; Lujewa 1987; Heinemann/Viehweger 1991

⁴ Vgl. Heinemann/Viehweger 1991

⁵ Vgl. Sommerfeldt/Starke 1992: 266; Vater 1994: 134

⁶ Die Einschränkung der Untersuchung auf die semantisch-denotativen Einheiten des Textes bedeutet gleichzeitig, dass Formen komplementärer Referenz in Anlehnung an Halliday/Hassan (1976: 285) ausgeschlossen werden. Komplementäre Referenz liegt etwa zwischen *die Amseln* und *die Buchfinken* bei Bezugnahme auf das gemeinsame semantische Feld, die gemeinsame Gesamtmenge *die Vögel* vor.

recht festen Regeln folgt und daher mit viel Redundanz, die die Perzeption erleichtert, geladen ist; hier sind die formalen Einheiten bis auf einige Ausnahmen kompakt, d.h. bilden ununterbrochene Ketten von Zeichen. Der andere, semantische Strom, ist turbulent, chaotisch, schwer voraussagbar. Auch hier gibt es Einheiten, die jedoch nicht kompakt sind, sondern aus Teilen bestehen, die im Text verstreut sein können. Ihre Größe, Streuung und Folge erzeugen bestimmte Muster, die (stochastischen) Gesetzen gehorchen. Die Aufgabe einer Textanalyse, die mehr als die Summe der verwendeten Zeichen erfassen will, besteht aber gerade darin, diese Einheiten und ihre Bildungsmuster zu finden.

Auf Grund des Referenz-Charakters der denotativen Relationen wird angenommen, dass diese Denotationsklassen bilden (auf welche alle Wörter eines Textes verteilt werden können) die dieselbe Entität in der (inneren oder äußeren) Realität bezeichnen, ungeachtet der Tatsache, ob Rekurrenz, Substitution, Paraphrasierung oder andere Formen der lexikalischen Wiederaufnahme vorliegen. Die denotativen und daher semantischen Klassen werden im Rahmen der Denotativen Textanalyse als *Hrebs* bezeichnet. Sämtliche Wörter (oder auch Morpheme, Phrasen usw.), die einem Hreb zugeordnet werden können, bilden somit eine Denotationsklasse, d.h. eine spezielle Art linguistischer Einheit, deren Status weder grammatisch noch phonetisch, sondern rein semantisch-denotativ ist. Bezüglich der Zuordnung der Wörter eines Textes zu den jeweiligen Hrebs sind in Ziegler/Altmann (2002) und an anderer Stelle in Ziegler/Altmann (2003) sowie Ziegler (2003) eine Reihe vorläufiger und tentativer Regeln aufgestellt worden, die an dieser Stelle – in Anbetracht des zur Verfügung stehenden Rahmens – nicht gesondert angeführt werden.⁷

Für den hier darzustellenden Ansatz ist wesentlich, dass eine Etablierung der Hrebs als sprachliche Einheiten es erlaubt, einen Text hinsichtlich seiner denotativen Struktur, seiner Musterbildungen auf denotativer und mithin semantischer Grundlage, zu untersuchen. Es ist dabei anzunehmen, dass Muster dieser Art durch statistische Gesetze erfasst werden können. Für eine linguistische Untersuchung scheint dabei besonders die Frage von Interesse, inwiefern die Parameter der Gesetze für bestimmte Textsorten oder Kommunikationsformen oder spezifische Varietäten charakteristisch sind.⁸

⁷ Vgl. Ziegler/Altmann 2002; Ziegler/Altmann 2003

⁸ Zu dem Status der denotativ-semantischen Einheiten als sprachliche Entitäten vgl. Ziegler/Altmann 2002

2 Die Denotative Textanalyse

Denotative Relationen erzeugen also Denotations- oder Wort-Hrebs, die aus allen Wörtern des Textes bestehen, die dieselbe Entität in der (äußeren oder inneren) Realität bezeichnen, ungeachtet der Tatsache ob Rekurrenz, Substitution oder andere Formen der lexikalischen Wiederaufnahme vorliegen.⁹ Diese Bezeichnung kann allgemein oder rein textspezifisch sein. Ein Wort-Hreb hat daher immer nur ein einziges Denotat ohne Rücksicht darauf, wie viele wortähnliche Ausdrücke er enthält. Die Wörter des Hrebs bilden eine Denotationsklasse, d.h. eine spezielle Art linguistischer Einheit, deren Status weder grammatisch noch phonetisch, sondern rein semantisch-denotativ ist. Mit der Etablierung des Denotations-Hrebs können zahlreiche formale Probleme umgangen werden. Es wird nicht mehr gefragt, ob die Pronomina *er*, *sie*, *es* ein oder drei Lexeme bilden. Alle drei gehören mit Sicherheit zu unterschiedlichen Hrebs. Allein das Personalpronomen *er* kann ja im Text auf ganz verschiedene Entitäten verweisen. Grundsätzlich wird also mit dem vorliegenden Verfahren die von J. S. Mill 1843 eingeführte Dichotomie von *Denotation* und *Konnotation* aufgegriffen.¹⁰

Um einen Eindruck der methodologischen Vorgehensweise der Denotativen Textanalyse an dieser Stelle zu ermöglichen, wird im Weiteren die Vorgehensweise exemplarisch anhand des Textes „Erlkönig“ von Johann Wolfgang von Goethe dargestellt.¹¹

2.1 Die Etablierung der denotativ-semantischen Einheiten

Bereits in Ziegler/Altmann (2002) und an verschiedenen anderen Stellen sind einige vorläufige und tentative Kriterien zur Etablierung der Hrebs angegeben worden, die daher hier nicht vollständig wiederholt werden sollen. Exemplarisch seien lediglich folgende Kriterien angeführt.

- Eine wortähnliche Einheit an einer bestimmten Stelle (Position) des Textes gehört grundsätzlich nur zu einem Hreb; jedoch kann eine identische Einheit an einer Stelle zu einem, an einer anderen zu einem zweiten Hreb gehören, z.B. in den Phrasen des „Erlkönigs“ *mein* Vater, *mein* Sohn, *meine* Töchter gehört *mein* im

⁹ Vgl. Linke/Nussbaumer/Portmann 1996: 215ff.

¹⁰ Mill verstand unter der Denotation noch die Extension der Bedeutung, den Bezug auf die Gesamtmenge der bezeichneten Gegenstände, und setzte den Begriff der Konnotation mit der Intension einer Bedeutung gleich. Er bezeichnete also jene Eigenschaften als Konnotationen, die es erlauben, Individuen als Elemente einer Klasse zu erkennen, sie unter einen Begriff zu subsumieren; vgl. Mill (1843), Winkler (1989).

¹¹ Der Text ist vollständig als Anhang zu diesem Beitrag abgedruckt; vgl. Hamburger Ausgabe 1998, Bd. 1: 154f.

ersten Fall zum Sohn-Hreb, im zweiten zum Vater-Hreb, im dritten zum Erbkönig-Hreb. Im Allgemeinen beziehen sich Pronomina auf unterschiedliche Entitäten der Realität.

- Konjugations- und Flexionsendungen ergeben im Deutschen nur einen grammatischen Bezug im Rahmen des Satzes. Diese Regel kann in anderen Sprachen aber ganz anders ausfallen und muss individuell entschieden werden. Auch über die Zuordnung von Possessivpronomina muss je nach Sprache entschieden werden. Im Ungarischen könnten die Wörter *házam* 'mein Haus' oder *dolgozok* 'ich arbeite' jeweils zu zwei unterschiedlichen Hrebs gehören, weil hier die Possessivität und Person mit einem Affix hinreichend ausgedrückt werden. Im Ungarischen ist das selbständige Pronomen in solchen Fällen fakultativ (*én dolgozok*, *én házam*), im Deutschen ist es obligatorisch (*ich arbeite*, *mein Haus*). Im hier dargestellten Verfahren werden die Personalendungen ignoriert, da eine morphematische Analyse nicht intendiert ist.
- Synonyme gehören im Allgemeinen zum gleichen Hreb, Homonyme nicht; polyseme Wörter können unterschiedlichen Hrebs zugeschrieben oder in einem belassen werden. So kann *ist* bedeuten: 'existieren', 'Eigenschaft haben', 'sich an einem Ort befinden', 'identisch sein mit' usw.

Wie sich in den vorangegangenen Beispielen andeutet, sind die Kriterien für die jeweils zu untersuchende Sprache spezifisch zu definieren. Da eine Hrebetablierung nach den konstatierten Kriterien bisher (noch) nicht automatisch durchgeführt werden kann, sondern auf eine individuelle Zuordnung und Disambiguierung zurückgegriffen werden muss, liegt hier gleichzeitig ein Schwachpunkt der Analyse. Auch bei sehr klar definierten Kriterien besteht z.B. aufgrund kontroverser grammatischer Modelle oder der Tatsache, dass die Schnittstelle zwischen Semantik und Denotation in der Linguistik sehr unterschiedlich erfasst wird, die Möglichkeit, dass ad hoc Entscheidungen getroffen werden müssen. Dennoch liefern gerade die Kriterien die notwendigen Anfangsbedingungen, unter denen die Analyse Gültigkeit erlangen kann.

Für eine Ermittlung der denotativ-semantischen Einheiten (Hrebs) des Textes werden nun alle Wörter nach ihrer Position im Text nummeriert. Für die ersten zwei Zeilen des Erbkönigs ergibt sich also

1 2 3 4 5 6 7 8
Wer reitet so spät durch Nacht und Wind

9 10 11 12 13 14 15
Es ist der Vater mit seinem Kind

In einem nächsten Schritt werden anschließend alle Wörter, die dasselbe denotieren ermittelt und zu Hrebs zusammengefasst. Für den hier analysierten Text resultiert z.B. als erster Hreb

(wer 1, Vater 12, seinem 14, er 16, er 24, er 28, mein 32, Vater 42, du 43, mein 53, Vater 87, Vater 89, du 92, mein 103, Vater 139, Vater 141, du 144, mein 152, mein 154, ich 156, Vater 185, Vater 187, Vater 200, er 202, er 205, seinen 220),

als zweiter

(reitet 2, reitet 203),

als dritter

(so 3, so 38, so 165, so 180) usw.

Die Zahlen hinter den Wörtern eines Hrebs geben die Position im Text wider. Insgesamt lassen sich 96 verschiedene denotativ-semantische Klassen (Hrebs) für den „Erlkönig“ ermitteln. Mittels der automatischen Anpassung über die Altmann-Fitter-Software konnte dabei gezeigt werden, dass die Rangordnung der Hrebs einer Wahrscheinlichkeitsverteilung (Zipf-Mandelbrot und Zipf-Alekseev) folgt.¹²

Die überwiegende Mehrheit der erwähnten 96 Hreb-Klassen enthält lediglich ein Element/Wort (69), d.h. in diesen Fällen ist keine semantische Referenz auf der Textebene zu konstatieren ist. Den **denotativ-semantischen Kern** des Textes bilden die Klassen, die die meisten Lexeme, jedoch mindestens zwei enthalten. Dies sind für den „Erlkönig“:

| | |
|----------------|------------|
| {Kind} mit | 8 Lexemen |
| {Vater} mit | 7 Lexemen |
| {Erlkönig} mit | 4 Lexemen. |

Für die Zugehörigkeit zum Kern des Textes ist also nicht die absolute Häufigkeit des Hrebs, sondern die lexikalische Variation entscheidend. Ansonsten würden in langen Texten vor allem Hilfsörter und Synsemantika zum Kern gehören.

Den Kern kann man nach Bedarf einschränken, für die hier vorgenommene Analyse werden im Kern die Hrebs erfasst, deren Lexemumfang $|\{H_i\}| \geq 2$ ist. Alle Hrebs, die nicht zum Kern gehören, bezeichnen wir als **Peripherie** des Textes. Die Summe dieser Zahlen ist wie folgt zu verzeichnen:

$$|Kern| = \sum_{i \in Kern} |\{H_i\}|$$

¹² Vgl. Altmann-Fitter 1997; Ziegler/Altmann 2002; nähere Informationen zur Software sind erhältlich unter <http://www.ram-verlag.de>

2.2 Thema und Topikalität im Text

Mit den ermittelten denotativ-semantischen Einheiten lassen sich im Rahmen der Denotativen Textanalyse eine Reihe von Operationen durchführen, die es erlauben verschiedene Textmaße – etwa hinsichtlich der Textkompaktheit, der Diffusität des Textes, der Topikalität oder der rhematischen Schichtung des Textes – zu berechnen. Von den vielfältigen Möglichkeiten soll hier lediglich die Charakterisierung der Topikalität des Textes exemplarisch angeführt werden.

Obwohl wir von unserem vorlinguistischen Alltagsverständnis her in der Regel keine Schwierigkeit haben, das Thema eines Textes zu bestimmen, ist es der Textlinguistik bis heute nicht gelungen, ein operationalisiertes Verfahren zur Bestimmung eines Textthemas vorzulegen, d.h. es ist nicht gelungen, das Thema eines Textes mit rein linguistischen Methoden eindeutig zu ermitteln.¹³ Mit der denotativen Textanalyse ist versucht worden, ein solches Verfahren anzubieten,

Im Anschluss an Agricola wird unter dem Thema eines Textes sein „begrifflicher Kern“ verstanden, dies aber auf denotativer Grundlage.¹⁴ Unter dem Begriff *Thema* werden zentrale Kerninformationen erfasst, die die semantischen (denotativen) Informationen des Textes dominieren.¹⁵ Auf der einen Seite erscheint somit das Thema eines Textes als semantisches Destillat, welches aus dem konkreten Text mittels der denotativen Textanalyse gewonnen werden kann; auf der anderen Seite ist das Thema des Textes „[...] ein Informationspotential, das einen Erwartungsrahmen schafft, auf den hin einzelne Textbestandteile interpretiert werden und der somit konstitutiv für die Kohärenz ist“.¹⁶ In Anlehnung an van Dijks Konzept des *Themawortes* wird daher im Weiteren auch von thematischen Hrebs gesprochen, da durch die in einem thematischen Hreb versammelten Wörter gleicher Denotation das Thema explizit ausgedrückt werden kann und ihnen damit eine wesentliche kognitive Funktion für eine Interpretation des Textes zukommt.¹⁷

Insofern sind die Kern-Hrebs des Textes die textuellen Einheiten, die generell in der Lage scheinen, das Thema eines Textes auszudrücken. Um zu ermitteln, inwieweit sie von dieser Möglichkeit Gebrauch machen und tatsächlich thematische Information tragen und um gleichzeitig ein Maß zur Differenzierung der verschiedenen Kern-Hrebs anzubieten, ist die **Topikalität** eines Kern-Hrebs zu berechnen.

Bezeichnet man mit $|\{H_i\}|$ den Umfang (= lexikalische Variation – gezählt in Anzahl der Lexeme) des Hrebs i , dann ergibt sich für die Topikalität eines Kern-Hrebs das Maß

¹³ Vgl. Linke/Nussbaumer/Portmann 1996.

¹⁴ Vgl. Agricola 1979.

¹⁵ Vgl. Kallmeyer 1980.

¹⁶ Kallmeyer 1980: 254.

¹⁷ Vgl. van Dijk 1980: 50.

$$(2.1) \quad T(Hreb_i) = \frac{|\{H_i\}|}{\sum_{i \in Kern} |\{H_i\}|} = \frac{|\{H_i\}|}{|Kern|},$$

d.h. der Umfang des Hrebs H_i , für den $|\{H_i\}| \geq 2$ gilt, dividiert durch die Summe der Umfänge aller Kern-Hrebs. Zuvor ist festgestellt worden, dass es im „Erlkönig“ drei Kern-Hrebs mit $|Kern| = 8 + 7 + 4 = 19$ gibt, daher gilt für den

$$\begin{aligned} T_{(Kind)} &= |Kind_M|/|Kern| = 8/19 = 0.42 \\ T_{(Vater)} &= |Vater_M|/19 = 7/19 = 0.37 \\ T_{(Erlkönig)} &= 4/19 = 0.21. \end{aligned}$$

Mit dem Maß

$$(2.2) \quad TK = \frac{|Kern|}{n},$$

wobei n die Zahl der Hrebs im Text ist, drückt die **Konzentriertheit des Kerns** aus und besteht aus der Summe der Umfänge der Hrebs im Kern, dividiert durch die Zahl aller Hrebs im Text. Auf diese Weise ergibt sich für den „Erlkönig“

$$TK_{(Erlkönig)} = 19/96 = 0.20.$$

Es ist davon auszugehen, dass mit der Analyse der thematischen Hrebs hinsichtlich Topikalität und Konzentriertheit des Kerns Verfahren zur Verfügung gestellt werden, die den Anforderungen an Operationalisierbarkeit genügen und es erlauben, Texte bezüglich thematischer Kriterien zu charakterisieren und zu vergleichen.

Während die Denotative Textanalyse und die im Vorangegangenen exemplarisch angeführten Textmaße allerdings lediglich einen Blick in die Prägung des Einzeltextes ermöglichen, kann eine Anwendung des Verfahrens an größeren Korpora erste Einblicke in die Phrasenbildung liefern und insofern den assoziativen Charakter einer Sprache offenbaren. Um dazu die Voraussetzungen zu schaffen, werden im Rahmen des hier gewählten Ansatzes Koinzidenzen berechnet und die Grafentheorie als geeignete Grundlage angenommen.

3 Koinzidenzen und Kohäsion – Grafentheoretische Grundlagen

Unter **Koinzidenz** wird ein definierter Berührungspunkt, d.h. eine Beziehung zwischen zwei Hrebs verstanden. Grundsätzlich korrespondiert eine solche Auffassung von Koinzidenz dabei mit dem textlinguistisch geprägten Begriff der **Kohäsion**, ist

aber demgegenüber statistisch bestimmt.¹⁸ Dennoch verweist die Koinzidenz auf kohäsiv verbundene Elemente eines Textes und ist in ihrer Gesamtheit (alle Koinzidenzen des Textes) auch als **Maß der Kohäsion** eines Textes unter den jeweiligen Untersuchungsaspekten zu verstehen. Selbstverständlich können zahlreiche Koinzidenzen definiert werden. Im Folgenden wird aber nur eine Art exemplarisch dargestellt.

Elemente zweier Hrebs können als grammatisch koinzident (grammatisch kohäsiv) bezeichnet werden, wenn sie eine grammatische Konstruktion bilden, d.h. grammatisch verbunden sind. Eine lockerere Definition würde zulassen, dass es reicht, wenn sie in einer grammatischen Konstruktion gemeinsam vorkommen. Die erste Definition wäre allerdings zu sprachspezifisch, denn in einer Sprache können zwei Wörter durch Kongruenz verbunden sein, in einer anderen, wo es keine Kongruenz gibt, nicht. Es ist daher auf eine breitere Definition zurückzugreifen.

Die semantische Koinzidenz (semantische Kohäsion) kann man sich als gemeinsames Vorkommen einer Einheit eines Textes in einem semantischen Feld vorstellen, aber dazu müsste zunächst die ganze Sprache bezüglich ihrer semantischen Felder analysiert werden, oder nur sehr lange Texte dürften untersucht werden. Weiter müsste gezeigt werden können, dass im gegebenen Text ein semantisches Feld aufgespannt ist, wozu natürlich ein statistischer Test nötig wäre.

Da ein Begriff der Koinzidenz gesucht wird, der ohne zusätzliche Kriterien in jeder Sprache und für jeden Text anzuwenden ist, wird die **positionale Koinzidenz** stellvertretend für die oben genannten Koinzidenzen gewählt. Dementsprechend sind zwei Hrebs koinzident, wenn sich zeigen lässt, dass ihre Elemente in einem (syntaktischen, semantischen, rhythmischen usw.) Rahmen signifikant oft zusammen vorkommen. Als syntaktisch-semantischen Rahmen kann man den Satz, als rhythmischen den Vers oder die Strophe annehmen (andere Möglichkeiten sind nicht ausgeschlossen).

Gemeinsames Vorkommen in dem gegebenen Rahmen ist für die Koinzidenz eine notwendige Bedingung: Es reicht nicht, wenn zwei Wörter Kongruenz aufweisen – z.B. beide im Akkusativ sind – aber zehn Sätze voneinander getrennt stehen; es reicht nicht, wenn sie zum gleichen semantischen Feld gehören, aber voneinander getrennt verwendet werden und nicht einmal das gleiche Denotat haben. Und es genügt auch nicht, wenn sie im vorbestimmten Rahmen zusammen vorkommen, jedoch rein zufällig. Die Signifikanz der Kookkurrenz muss nachgewiesen werden. Sie gilt gleichzeitig als hinreichende Bedingung für jegliche Art der Koinzidenz, ohne sie weiter spezifizieren zu müssen.

¹⁸ Vgl. Linke/Nussbaumer/Portmann 1996.

Um die positionale Koinzidenz untersuchen zu können, ist folgendermaßen zu verfahren. Zunächst wird der gesamte Text entsprechend der Hrebzugehörigkeit seiner Wörter numerisch kodiert. Das kodierte Gedicht erhält dann folgende Form:

| | | | |
|---|----------------------|----------------------|----------------------|
| | 2,17,8,27,28,29,5,1 | 1,39,1,40,23,7,3 | 3,16,70,71,72 |
| 8 | 9,6,4,2,7,2,1 | 41,15,42,43,3,7,1 | 2,1,2,1,2,14,9,73 |
| | 2,12,4,1,30,10,4,13 | 24,44,45,6,46,4,47 | |
| | 2,19,1,31,2,20,1,32 | 3,48,12,24,49,50 | 9,74,75,76,77,8,78 |
| | | 1,2,1,2,5,51,2,11 | 3,79,1,3,80,1,15,81 |
| | 2,1,21,33,1,8,34,1,3 | 21,3,1,52,53 | 5,6,1,11,82,8,83,3,8 |
| 5 | 14,2,2,4,3,11 | 6,25,54,25,2,1 | 4 |
| | 4,3,7,36,5,37 | 10,55,56,57,4,18 | 1,2,1,2,85,19,3,1,86 |
| | 2,1,9,6,22,38 | 58,59,1,1,7,3,23 | 3,12,1,22,87,88 |
| | | 3,16,60,1,61,15 | 4,2,89,2,17,90 |
| | | 3,16,62,4,63,64 | 2,20,10,13,26,91,1 |
| | | 5,65,5,66,5,67,1,68 | 92,4,93,7,94,5,95 |
| | | | 10,2,13,26,1,6,96 |
| | | 1,2,1,2,5,14,2,11,69 | |

Diese Darstellung eignet sich dazu, die Koinzidenzstruktur des Textes zu untersuchen. Mit N ist die Zahl der Verse (Sätze, Absätze oder anderer Rahmeneinheiten) des Textes bezeichnet, die Zahl der Verse, in denen Hreb A vorkommt mit M , die Zahl der Verse, in denen Hreb B vorkommt mit n und die Zahl der Verse, in denen die beiden gemeinsam vorkommen (mehrfaches Vorkommen eines Hreb in einem Vers ist irrelevant) mit x , dann ist die Wahrscheinlichkeit, dass A und B mindestens in x Versen zusammen vorkommen

$$(3.1) \quad P(X \geq x) = \sum_{j=x}^{\min(M,n)} \frac{\binom{M}{j} \binom{N-M}{n-j}}{\binom{N}{n}},$$

d.h. als die Summe einzelner hypergeometrischer Wahrscheinlichkeiten zu erfassen. Bei sehr langen Texten kann man zu der Poisson-Verteilung oder auch zur Normalverteilung übergehen.¹ Je kleiner diese Wahrscheinlichkeit, desto stärker die positionale Koinzidenz der Hrebs.

¹ Vgl. Altmann 1988.

Als Beispiel wird die Koinzidenz von (Vater) und (Kind) aus dem „Erlkönig“ berechnet. Es gibt im „Erlkönig“ $N = 32$ Verse, Elemente aus (Vater) kommen in $M = 15$ Versen vor, Elemente aus (Kind) in $n = 21$ Versen und gemeinsam kommen sie in 12 Versen vor. Setzen wir diese Zahlen in (3.1) ein, so resultiert

$$P(X \geq 12) = \frac{\binom{15}{12} \binom{32-15}{21-12}}{\binom{32}{21}} + \frac{\binom{15}{13} \binom{32-15}{21-13}}{\binom{32}{21}} + \frac{\binom{15}{14} \binom{32-15}{21-14}}{\binom{32}{21}} + \frac{\binom{15}{15} \binom{32-15}{21-15}}{\binom{32}{21}} = 0.0857 + 0.0198 + 0.0023 + 0.0001 = 0.1079.$$

Mit diesem Verfahren müssen nun alle Hrebpaaare getestet werden. Für so umfangreiche Daten wie die Hrebs eines Textes lässt sich diese Aufgabe nur mit einer geeigneten Software bewältigen. Das Programm **ADJAZENZ**, das extra für die Denotative Textanalyse von R. Köhler entwickelt wurde, erledigt diese Aufgabe sehr schnell und vor allem ohne Fehler.²

In einem folgenden Schritt kann jetzt eine Ebene α gewählt und als koinzident alle Hrebs betrachtet werden, bei denen $P(X \geq x_c) \leq \alpha$. Zeichnet man die Hrebs als Punkte (Ecken) in eine Ebene und verbindet diejenigen, bei denen $P \leq \alpha$, dann erhält man einen Grafen, der als α -Graf bezeichnet werden kann (vgl. Abb. 1).

Es ist zu bedenken, dass α eine Konvention ist. Je kleiner α , desto weniger Zusammenhänge werden zu registrieren sein. Hier ist $\alpha = 0.1$ gewählt, was etwas höher als die üblichen Grenzen in der Statistik ist. Die angegebene Koinzidenz ist im Grunde *ungewichtet*, weil man sie nur bei der Überschreitung einer Schwelle als existent betrachtet. Man könnte die Kanten des Grafen auch *gewichten*, wobei die Gewichtung die Stärke der Assoziation zweier Hrebs darstellen würde. Am einfachsten wäre dies über die berechnete Wahrscheinlichkeit oder ihr Komplement, d.h. $1 - P(X \geq x_c)$ zu leisten. Dies würde aber einen Grafen mit viel mehr Kanten ergeben, er wäre unübersichtlicher und schwerer handhabbar. Bei sehr kurzen Texten – kurz im Sinne der Satzanzahl, nicht der Wortanzahl – wird man diese Gewichtung in Betracht ziehen müssen. Dies ist eine Möglichkeit, auch größere gewichtete Grafen für die Textanalyse zu verwenden.

² Vgl. Ziegler/Altmann 2002; das Programm ist als Windows-Anwendung auf einem Datenträger beigelegt.

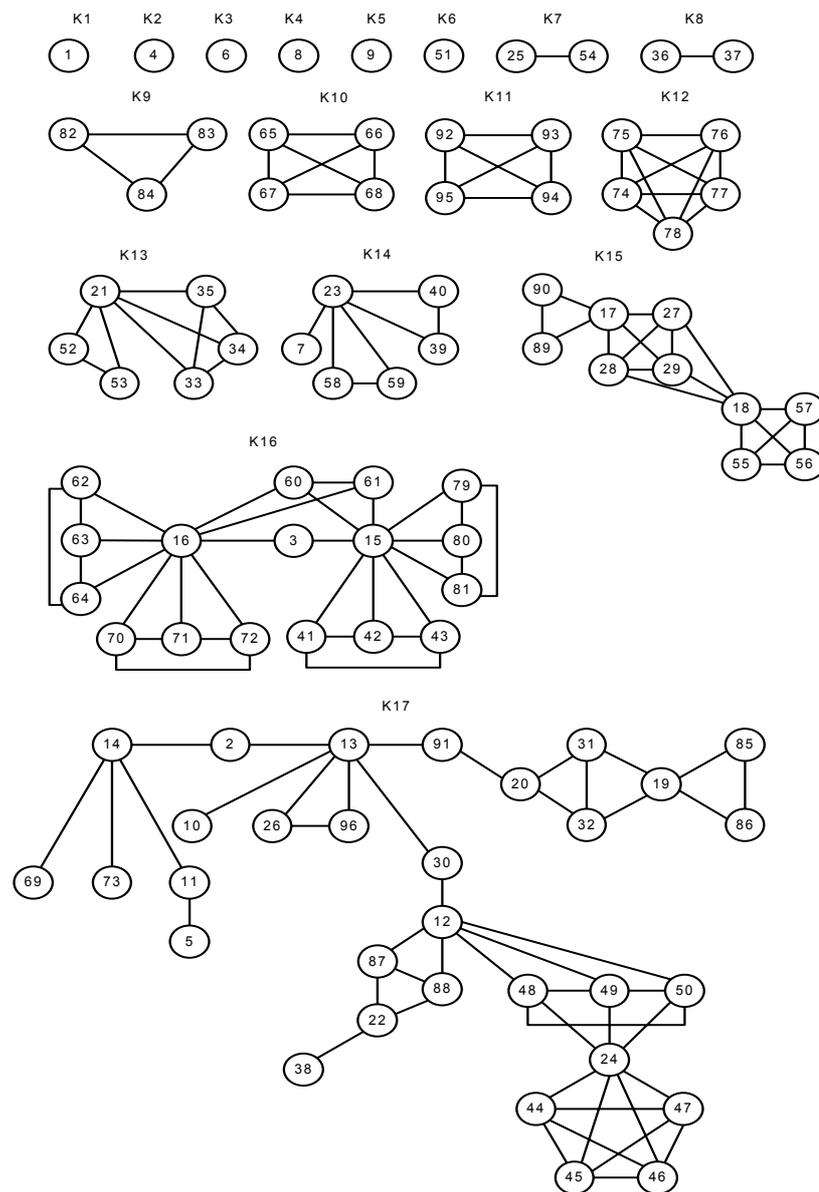


Abb. 1: Graf der Koinzidenzen im Erbkönig.

4 Assoziativität und konnotative Struktur

Im vorigen Abschnitt ist aus den Koinzidenzen ein Graf erstellt worden, der die signifikanten (nicht-zufälligen) Zusammenhänge des Textes abbildet. Bestimmte Aspekte der Grafentheorie erlauben es aber, die Textkohärenz nicht nur visuell darzustellen, sondern auch mit Maßen zu charakterisieren. Wie man die Kohärenz eines Textes auch immer definiert, die folgenden Methoden lassen sich immer anwenden.

Ein Graf besteht aus Ecken (Knoten) und aus Kanten, die die Ecken miteinander verbinden. Hier haben wir es mit ungerichteten Grafen zu tun, bei denen die Richtung der Verbindung nicht angegeben wird. Ein Textgraf ließe sich aber *cum grano salis* auch als ein gerichteter Graf darstellen.

Zwei Ecken sind *adjazent*, wenn sie mit einer Kante verbunden sind, und man sagt, dass eine Ecke mit einer Kante *inzidiert*. Die oben benutzte statistische Koinzidenz ist etwas anderes. Eine Ecke, die mit keiner Kante inzidiert, heißt *isoliert*. In den obigen Grafen sieht man mehrere isolierte Ecken. Die Zahl der Ecken $|E| = n$ heißt *Ordnung* des Grafen, die Zahl der Kanten $|K| = m$ *Größe* des Grafen.

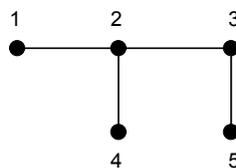


Abb. 2: Beispielfraf.

Die Zahl der Kanten, die mit einer Ecke v inzidieren, heißt *Eckengrad* $d(v)$. Wie in der Abbildung 2 zu sehen ist, haben die Ecken dieses Grafen die Grade $d(v_1) = d(v_3) = d(v_5) = 1$, $d(v_2) = 3$, $d(v_4) = 2$.

Der Eckengrad eines Hrebs verweist auf die Assoziativität des Hrebs mit anderen. Während ein Hreb selbst denotativ definiert ist, drückt seine Verbindung zu anderen Hrebs seine potentielle **Assoziativität** und/oder seinen **konnotativen Status** aus. Diese können lokal sein, d.h. nur auf den gegebenen Text beschränkt, oder sie deuten an, dass es eine Tendenz gibt, dass mit dem Hreb komplexere Begriffe oder Phrasen gebildet werden können. Die Assoziativität ist üblicherweise gering, wenn der Hreb umfangreich ist, d.h. wenn viele unterschiedliche Formen (Lexeme) zu beobachten sind. Die folgende Tabelle gibt einen Überblick zu den Graden der Hrebs des hier gewählten Beispieltexes (vgl. Tab. 1).

Tabelle 1
Grade der Hrebs im „Erlkönig“

| Grad | Hrebs | Anzahl |
|------|---|--------|
| 0 | {Kind, der, ist, so, es, hörest} | 6 |
| 1 | {und, mit, in, ruhig, Kron, Schweif, Nebelstreif, bleibe, dort, genau} | 10 |
| 2 | {Vater, Erlkönig, nicht, das, wohl, liebes, komm, leise, verspricht, willst, feiner, willig, brauch, Gewalt, jetzt, an (Adv.), grausets, geschwind, ächzende, tot} | 20 |
| 3 | {halten, ein (Art.), sicher, warm, birgst, bang, Gesicht, gar, Spiele, spiel, dürrer, Blättern, säuselt, sollen, warten, führen, nächtlichen, Reihn, wiegen, tanzen, singen, ein (Adv.), am, düstern, Ort, liebe, reizt, Gestalt, Leids, getan, erreicht, Hof, Mühe, Not} | 34 |
| 4 | {sehen, fassen, spät, durch, Nacht, bunte, Blumen, an, Strand, Mutter, gülden, Gewand, scheinen, die (Pl. fem.), alten, Weiden, grau} | 17 |
| 5 | {reiten, was, gehen} | 3 |
| 6 | {hat, Arm, Wind} | 3 |
| 7 | {manch} | 1 |
| 9 | {schön, Tochter} | 2 |

Da bei der Gradrechnung jede Kante zweimal gezählt wird, ist der Grad des ganzen Grafen $d(G) = 2m$, d.h. er entspricht der doppelten Zahl aller Kanten.

Die Klassifikation der Hrebs nach Graden ist eine assoziative (konnotative, phrasologische) Klassifikation. Hrebs mit gleichem Grad haben semantisch nichts Gemeinsames, der Grad zeigt lediglich ihre konnotative Ausprägung, ihre assoziative Potenz.

Für unsere Zwecke eignet sich die Charakterisierung des Textes durch den *relativen Gesamtgrad* des Grafen, $d_{rel}(G)$, den wir so berechnen, dass wir die Zahl der Kanten, m , durch die maximale Zahl der Kanten dividieren. Wenn jede der n Ecken mit jeder der restlichen $n-1$ Ecken adjazent wäre, so hätten wir insgesamt $n(n-1)/2$ Kanten, so dass wir

$$(4.1) \quad d_{rel}(G) = \frac{2m}{n(n-1)}$$

bekämen. Den *durchschnittlichen Eckengrad* pro Hreb, $\bar{d}(G)$, bekämen wir als

$$(4.2) \quad \bar{d}(G) = \frac{2m}{n}.$$

Und schließlich ist

$$(4.3) \quad \Delta(G) = d_{\max}(v).$$

der *maximale Eckengrad* im Graf. Für den Erlkönig bekommt man – wie mit $m = 139$, $n = 96$ leicht zu berechnen ist

$$\begin{aligned} d(G) &= 278, \\ d_{rel}(G) &= 278/[96(95)] = 0.0305 \\ \bar{d}(G) &= 278/96 = 2.90, \\ \Delta(G) &= 9 \end{aligned}$$

Diese Maße zeigen auch einen Aspekt der Stärke der Verbundenheit des Textes, sind aber keine Maße der Einbindung oder Zentralität einzelner Hrebs. Man kann sehen, dass z.B. im „Erlkönig“ die zentralen Hrebs {Kind, Vater, Erlkönig} sehr niedrige Eckengrade haben. Es braucht kaum erwähnt zu werden, dass eine andere Signifikanzgrenze andere Resultate geliefert hätte. Bei künstlerischen Texten ist generell eine stärkere Verbundenheit zu erwarten. Das Maß (4.1) kann aber als ein **Maß der konnotativen Konzentration** betrachtet werden.

5 Zusammenfassung

Bereits an anderer Stelle ist darauf aufmerksam gemacht worden, dass prinzipiell Viehwegers Definition der *nominativen Ketten* gefolgt wird und zwar insbesondere im Hinblick auf die Tatsache, dass den nominativen Ketten eine Isotopiekette mit denotativer Bedeutung entspricht.³ Diese Isotopiekette wird im Rahmen unseres Ansatzes durch die im Vorangegangenen etablierten Hrebs fundamntiert. Bereits Lujewa (1987) geht davon aus, dass eine **konnotative Textpotenz** auf der Ebene der nominativen Ketten überprüft werden kann. Sie definiert die konnotative Textpotenz als eine textuelle Implikation, die nicht subjektive Interpretation, sondern vielmehr textbedingte Erscheinung ist und zwangsläufig zur Bildung eines Subtextes führt. Indem wir die eine Untersuchung der Hrebs, als denotative Entitäten des Textes, nach ihren Graden durchführen, schließen wir uns dieser Meinung an und ermitteln analog zu der postulierten konnotativen Textpotenz an dieser Stelle die **konnotative Konzentration** des

³ Vgl. Viehweger 1978; Lujewa 1987; Ziegler/Altmann 2003.

Textes.⁴ Indem angenommen wird, dass eine konnotative Wirkung auch textuell verankert ist, befinden wir uns dabei genau an der Schnittstelle der Textanalyse, die auf das Anschaulichste verdeutlicht, dass eine semantisch-denotative Betrachtung, wie sie hier verstanden ist, auch immer im konkreten Zusammenhang mit der Pragmatik des Textes zu sehen ist. Natürlich muss der konnotative Status eines Hrebs erst durch den Leser/Hörer aktualisiert werden, um seine Wirkung zu entfalten. Da diese Aktualisierungen wieder im Rahmen eines semantischen Intervalls vollzogen werden, erscheint es selbstverständlich, dass ein und derselbe Text ganz unterschiedlich rezipiert und interpretiert werden kann, d.h. unterschiedliche Subtexte produzieren kann. Die Klassifikation der Hrebs nach ihren Graden bietet die Möglichkeit, Texte hinsichtlich ihrer **potentiellen Subtexte** zu differenzieren und somit Variabilität sowie Grenzen der Variabilität der Sinnerschließung eines Textes zu unterscheiden, da diese durch die Ermittlung der textinternen konnotativen Konzentration bestimmbar werden.⁵ Mit der textinternen konnotativen Textkonzentration werden also „[...] aussage- bzw. sinnkonstituierende Funktionsmerkmale, die durch den Bezug der Zeichen [Hrebs] aufeinander und auf die Ganzheit der Textgestalt [...] emotional-assoziative Bewusstseinsprozesse auslösen“ bezeichnet.⁶

Da gleichzeitig der assoziative Status der Hrebs bestimmt worden ist, sind durch die Integration der Grafentheorie in die Textanalyse darüber hinaus auch verschiedene Schlussfolgerungen für die Phraseologie und Parömiologie denkbar. Durch die Analysen wird nämlich deutlich, welche Hrebs generell einer potentiellen Phrasenbildung offen stehen, d.h. welche Phrasen geeignet scheinen einen Phraseologismus (im Zusammenhang mit den durch die Eckengrade verbundenen Hrebs) zu bilden. Langfristig würden damit – natürlich nach zahlreichen weiteren Untersuchungen – prädikative Aussagen hinsichtlich der assoziativen/phraseologischen Substanz einer Sprache sowie zu dem sprachvermittelten Weltbild, denkbar. Bezüglich eines sich kontinuierlich vollziehenden Sprachwandels wäre die Linguistik damit in der Lage, eine rein deskriptive Betrachtung sprachlicher Phänomene aufzugeben. Der praktische Nutzen für z.B. den Sprachunterricht und andere Bereiche der angewandten Linguistik wäre unübersehbar.

⁴ Vgl. Lujewa 1987: 127f.

⁵ Vgl. Lerchner 1984.

⁶ Lerchner 1984: 44.

Bisher liegen im Rahmen der Denotativen Textanalyse Untersuchungen zu Referenzbeziehungen in verschiedenen Sprachen und Textsorten vor. Erste Analysen – u.a. in Form von Magister- und Doktorarbeiten – widmen sich der deutschen, englischen, der slowakischen, ungarischen und der tschechischen Sprache. Es bleibt zu wünschen, dass die Verfahren in zahlreichen weiteren Untersuchungen und Sprachen voran getrieben werden und die hier skizzierten ersten Ansätze eine genauere Überprüfung und notwendige Modifikation erfahren.

Quellentext

Erlkönig

Wer reitet so spät durch Nacht und Wind?
Es ist der Vater mit seinem Kind;
Er hat den Knaben wohl in dem Arm,
Er fasst ihn sicher, er hält ihn warm.

„Mein Sohn, was birgst du so bang dein Gesicht!“
„Siehts, Vater, du den Erlkönig nicht?
Den Erlenkönig mit Kron und Schweif?“-
„Mein Sohn, es ist ein Nebelstreif.“-

„Du liebes Kind, komm, geh mit mit!
Gar schöne spiele spiel ich mit dir;
Manch bunte Blumen sind an dem Strand,
Meine Mutter hat manch gülden Gewand.“

„Mein Vater, mein Vater, und hörest du nicht,
Was Erlenkönig mir leise verspricht?“ –
„Sei ruhig, bleibe ruhig, mein Kind:
In dürren blättern säuselt der Wind.“-

„Willst, feiner Knabe, du mit mir gehen?
Meine Töchter sollen dich warten schön;
Meine Töchter führen den nächtlichen reihn
Und wiegen und tanzen und singen dich ein.“

„Mein Vater, mein Vater, und siehst du nicht dort
Erlkönigs Töchter am düstern Ort?“-
„Mein Sohn, mein Sohn, ich seh es genau:
Es scheinen die alten Weiden so grau.“-

„Ich liebe dich, mich reizt deine schöne Gestalt;
Und bist du nicht willig, so brauch ich Gewalt.“
„Mein Vater, mein Vater, jetzt fasst er mich an!
Erlkönig hat mir ein Leids getan!“-

Dem Vater grauset, er reitet geschwind,
Er hält in Armen das ächzende Kind,
Erreicht den Hot mit Mühe und Not:
In seinen Armen das Kind war tot.

Literatur

- Agricola, Erhard (1969): Semantische Relationen im Text und im System. Halle: Niemeyer.
- (1977): Text – Textaktanten – Informationskern. In: František Daneš, Dieter Viehweger (Hrsg.): *Probleme der Textgrammatik II*. Berlin: Akademie Verl., 11-32. (= *Studia grammatica XVIII*).
- (1979): Textstruktur, Textanalyse, Informationskern. 1. Aufl. Leipzig: Verl. Enzyklopaedie.
- Altmann Fitter (1997): Altmann-Fitter. Iterative fitting of probability distributions. Lüdenscheid: RAM-Verlag 1997. [Software].
- Altmann, Gabriel (1988): Wiederholungen in Texten. Bochum: Brockmeyer.
- Halliday, Michael A. K./Hasan, Ruqaiya (1976): *Cohesion in English*. London: Longman.
- HAMBURGER AUSGABE (1998) = Johann Wolfgang von Goethe. Werke. Hamburger Ausgabe in 14 Bänden. München: DTV.
- Heinemann, Wolfgang/Viehweger, Dieter (1991): Textlinguistik. Eine Einführung. Tübingen: Niemeyer. (= Reihe Germanistische Linguistik 115).
- Kallmeyer, Werner (1980): Einführung. 3. Aufl. Frankfurt a. M.: Athenäum. (= Lektürekolleg zur Textlinguistik 1).
- Köller, Wilhelm (1988): Philosophie der Grammatik. Vom Sinn grammatischen Wissens. Stuttgart: Metzler.
- Korzybski, Alfred (1973): *Science and Sanity. An Introduction to Non-Aristotelian Systems and General Semantics*. 4. Aufl. Lakeville/Conn.: International Non-Aristotelian Library Publ. Co.
- Lerchner, Gotthard (1984): Konnotative Textpotenz. *Beiträge zur Erforschung der deutschen Sprache* 4, 39.
- Linke, Angelika/Nussbaumer, Markus/Portmann, Paul R. (1996): Studienbuch Linguistik. 3., unveränd. Aufl. Tübingen: Niemeyer.
- Lujewa, Natalja (1987): Zur Erschließung der konnotativen Textpotenz auf der Ebene der nominativen Ketten. *Beiträge zur Erforschung der deutschen Sprache* 7, 117-126.
- Mill, John S. (1843): *A System of Logic*. London: Parker.
- Rapoport, Anatol (1972): Bedeutungslehre. Eine semantische Kritik. Darmstadt: Verl. Darmstädter Blätter.
- Rosengren, Inger (1980): Texttheorie. In: Hans Peter Althaus, Helmut Henne, Herbert Ernst Wiegand (Hrsg.): *Lexikon der Germanistischen Linguistik*. Studienausgabe. 2., vollst. neu bearb. u. erw. Aufl. Tübingen: Niemeyer, 275-286.
- Schmidt, Siegfried J. (1994): Kognitive Autonomie und soziale Orientierung. Konstruktivistische Bemerkungen zum Zusammenhang von Kognition, Kommunikation, Medien und Kultur. Frankfurt a.M.: Suhrkamp.
- Sommerfeld, Karl-Ernst/Starke, Günter (1992): Einführung in die Grammatik der deutschen Gegenwartssprache. 2., Neubearb. Aufl. Tübingen: Niemeyer.
- van Dijk, Teun A. (1977): *Text and Context. Explorations in the Semantics and Pragmatics of Discourse*. London: Longman. (= Longman Linguistics Library 21).
- (1980): *Macro-Structures. An Interdisciplinary Study of Global Structures in Discourse, Cognitions and Interaction*. Hillsdale/New Jersey.

- Vater, Heinz (1994): Einführung in die Textlinguistik. Struktur, Thema und Referenz in Texten. 2., überarb. Aufl. München: Fink.
- Viehweger, Dieter (1976): Semantische Merkmale und Textstruktur. In: Daneš František, Dieter Viehweger (Hrsg.): Probleme der Textgrammatik I. Berlin: Akademie Verl., 195-206. (= Studia grammatica XI).
- (1977): Zur semantischen Struktur des Textes. In: Daneš František, Dieter Viehweger (Hrsg.): Probleme der Textgrammatik II. Berlin: Akademie Verl., 103-117. (= Studia grammatica XVIII).
- (1978): Struktur und Funktion nominativer Ketten im Text. In: Wolfgang Motsch (Hrsg.): Kontexte der Grammatiktheorie. Berlin: Akademie Verl., 149-168. (= Studia Grammatica XVII).
- Ziegler, Arne (2001): Schreiber – Text – Leser. Denotative Textkonstituierung und diskurssemantische Gestalt in frühneuhochdeutschen Texten. In: Lenka Vaňková, Pavla Zajícová (Hrsg.): Aspekte der Textgestaltung. Referate der internationalen germanistischen Konferenz Ostrava 15.-16. Februar 2001. Ostrava, 197-210.
- (2003): Denotative Textanalyse. In: Gabriel Altmann, Reinhard Köhler, Raimund Piotrowski (Hrsg.): Quantitative Linguistik – Quantitative Linguistics. Ein internationales Handbuch. Berlin/New York: deGruyter, [erscheint].
- /Altmann, Gabriel (2002): Denotative Textanalyse. Ein textlinguistisches Arbeitsbuch. Wien: Edition Praesens.
- /Altmann, Gabriel (2003): Text Stratification. *Journal of Quantitative Linguistics*, [im Druck].