

Oliver Huber

Hyper-Text-Linguistik

TAH: Ein textlinguistisches Analysemodell für Hypertexte
Theoretisch und praktisch exemplifiziert am Problemfeld
der typisierten Links von Hypertexten im World Wide Web



Herbert Utz Verlag · Wissenschaft
München

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Zugleich: Dissertation, München, Univ., 2002

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, der Entnahme von Abbildungen, der Wiedergabe auf photomechanischem oder ähnlichem Wege und der Speicherung in Datenverarbeitungsanlagen bleiben – auch bei nur auszugsweiser Verwendung – vorbehalten.

Copyright © Herbert Utz Verlag GmbH 2003

ISBN 3-8316-0198-4

Printed in Germany

Herbert Utz Verlag GmbH, München

Tel.: 089/277791-00 – Fax: 089/277791-01

Komplettes Inhaltsverzeichnis

Prolog: Eine Bestandsaufnahme.....	1
1 Ausgangslage.....	1
1.1 Textlinguistische Motivation	1
1.2 Problemstellung.....	2
1.3 Vorgehensweise	3
1.4 Einschränkungen	4
1.5 Zielgruppe	5
2 Angaben zum Corpus	6
2.1 WWW als technische Plattform.....	6
2.2 Informationstexte als primäre Textsorte	6
2.3 Die Texte des Corpus'	8
3 Angaben zur Literaturlage.....	10
3.1 Literatur zu Hypertext	10
3.2 Literatur zur Textlinguistik	12
3.3 Literatur zu Textlinguistik und Hypertext	12
Teil A: Hypertext als Untersuchungsgegenstand.....	13
4 Hypertext: Ein historischer Abriss.....	14
4.1 Die Pioniere: Bush, Nelson und Engelbart	14
4.2 Der Durchbruch: Berners-Lee und das WWW	17
5 Text, E-Text, Hypertext und Hypertext-Netz.....	20
5.1 Abgrenzung Text vs. Teiltext, Hypertext vs. Knoten	20
5.2 Abgrenzung E-Text, Hypertext und Hypertext-Netz.....	22
6 Komponenten eines Hypertextes	24
6.1 Allgemein	24
6.2 Komponente I: Knoten	24
6.2.1 Struktur.....	25
6.2.2 Inhalt und Größe.....	26
6.2.3 Darstellungsform.....	27
6.2.4 Knoten-Typen.....	27
6.3 Komponente II: Links	28
6.3.1 Allgemeine Eigenschaften	29
6.3.2 Formale Unterscheidungsmöglichkeiten.....	30
6.3.3 Inhaltliche Unterscheidungsmöglichkeiten.....	32
6.3.4 Unterscheidung nach dem Travesierungsverhalten	33

6.3.5	Kritik der Link-Typologie.....	34
6.4	Navigations- und Orientierungskomponenten	35
6.4.1	Textsequenzierung, Navigation und Browsing.....	35
6.4.2	Orientierungsprobleme und kognitive Mehrbelastung	36
6.4.3	Traditionelle Orientierungshilfen.....	38
6.4.4	Verzeichnisse	39
6.4.5	Metainformationen	39
6.4.6	Typographische Auszeichnungen	40
6.4.7	Hypertext-spezifische Orientierungs- und Navigationshilfen	40
6.4.8	Browser-Software.....	40
6.4.9	Backtracking, Historie, Bookmarks	42
6.4.10	Graphische Übersichten	43
6.4.11	Anfrage- bzw. Suchmechanismen: Information Retrieval.....	43
6.4.12	Guided Tours.....	44
7	Zusammenfassung: Hypertexte im WWW.....	45
Teil B:	Ein Textlinguistisches Analysemodell für Hypertexte (TAH) ..	46
8	Textdefinitionen.....	47
8.1	<i>Text</i> als vorwissenschaftliche Größe.....	47
8.2	Text: Produkt und Produktion.....	47
8.3	Die Entwicklung des Terminus' <i>Text</i>	48
8.4	Exkurs: Einheiten zwischen Text und Satz.....	52
9	Textualitätskriterien	54
9.1	Kohäsion.....	55
9.2	Kohärenz	58
9.2.1	Kohärenz im Hypertext.....	60
9.2.2	Exkurs: Kohärenzerzeugung durch Schemata	60
9.3	Intentionalität	65
9.4	Akzeptabilität	66
9.5	Informativität.....	68
9.6	Situationalität	69
9.7	Intertextualität	70
9.8	Zusammenfassung.....	72
10	Dimensionen einer textlinguistischen Analyse.....	74
10.1	Allgemein.....	74
10.2	Kontextuelle Bedingungen.....	74
10.2.1	Kommunikationsform	75

10.2.2	Handlungsbereich.....	76
10.3	Textfunktion.....	77
10.4	Textstruktur.....	78
10.5	Textsorte.....	79
11	Paratextuelle Elemente.....	82
11.1	Traditionelle paratextuelle Elemente von Texten.....	82
11.2	Paratextuelle Elemente von Hypertexten.....	83
12	Makrostruktur und Makrostrukturanalyse.....	86
12.1	Allgemeines: Textthema und Makrostruktur.....	86
12.2	Exemplarische Makrostrukturanalyse eines Hypertextes.....	88
13	Referenzstruktur und referentielle Bewegung.....	92
13.1	Referenzbeziehungen in Texten.....	92
13.2	Modell der referentiellen Bewegung.....	93
13.3	Referentielle Bewegung in Hypertexten.....	94
13.4	Die Rolle der Quaestio.....	98
14	Zusammenfassung: Das Analysemodell TAH.....	100
14.1	Ausrichtung und Vorteile des Analysemodells.....	100
14.2	Einschränkungen.....	101
14.3	Überblicksdarstellung von TAH.....	103
14.4	Anmerkungen zu den Untersuchungsebenen.....	105
14.4.1	Anmerkungen zur Ebene <i>Hypertext</i>	105
14.4.2	Anmerkungen zu den Ebenen <i>Knoten</i> , <i>Absatz</i> und <i>Satz</i>	105
14.4.3	Anmerkungen zur Ebene <i>Link</i>	106
Teil C:	Fallbeispiel – Typisierte Links.....	108
15	Technische Voraussetzungen: Links im WWW.....	110
15.1	Datenhaltung, Zwischenschicht und Oberflächenstruktur.....	110
15.1.1	Von HTML zum Semantic Web.....	110
15.1.2	Die Ebenen einer Link-Typisierung.....	111
15.1.3	Die Oberflächenstruktur als Problembereich der Untersuchung ..	115
15.2	Linking in HTML und XML.....	117
15.2.1	HTML als SGML-Anwendung.....	118
15.2.2	HTML als XML-Anwendung.....	124
15.2.3	Das Hyperlink-Modell von XML: Xlink.....	129
15.2.4	Implementierung typisierter Links im Corpus'.....	130
15.2.5	Mögliche zukünftige Implementierungen.....	132

16	Vokabularien typisierter Links.....	138
16.1	Problemstellung.....	138
16.2	Linguistische Ansätze	139
16.2.1	Relation Element Theory.....	139
16.2.2	Rhetorical Structure Theory	145
16.3	Hypertext-spezifische Ansätze.....	154
16.3.1	Allgemeine Anmerkungen und kleine Pilotstudie	154
16.3.2	Typisierte Links in älteren Hypertext-Systemen.....	157
16.3.3	Typisierte Links in HTML	161
16.4	Zusammenfassung: Geeignete Vokabularien	175
17	Analyse des Corpus' mit Hilfe von TAH.....	176
17.1	Vorbemerkung zur Analyse	176
17.2	Allgemeine, exemplarische Analyse des Corpus'	177
17.2.1	Kontextuelle Aspekte	177
17.2.2	Kommunikativ-funktionale Aspekte.....	179
17.2.3	Konventionelle Aspekte	182
17.2.4	Strukturelle und paratextuelle Aspekte	182
17.2.5	Intertextuelle Aspekte	190
17.3	Analyse der Typisierten Links	194
17.3.1	Analyse der proprietären Codierung von SELF.....	194
17.3.2	Analyse der proprietären Codierung von SELF8.....	203
17.3.3	Analyse der proprietären Codierung von GRAMMIS.....	206
17.3.4	Analyse der proprietären Codierung von SPORT1	211
17.3.5	Versuch einer RST-Codierung des Corpus'	216
17.3.6	Versuch einer Codierung des Corpus' mit Hilfe des Dublin Core.....	218
17.3.7	Versuch einer Codierung des Corpus' mit Hilfe des <i>rel</i> -Attributs.....	219
17.4	Zusammenfassung: Auswertung der Ergebnisse	220
17.4.1	Aussagen zur Methode	220
17.4.2	Aussagen zu inhaltlichen Fragestellungen bzgl. der Ebene <i>Link</i>	222
Epilog:	Einsatzmöglichkeiten des TAH.....	230
Anhang:	Literatur und Verzeichnisse.....	233
A	Glossar	234
B	Fragebogen der Pilotstudie.....	238
C	Corpusverzeichnis	239
D	Literaturverzeichnis.....	240

Prolog: Eine Bestandsaufnahme

1 Ausgangslage

Die technische Entwicklung der Informationstechnologien schreitet mit großer Geschwindigkeit voran. Während Begriffe wie INTERNET, HTML und WWW vor fünf Jahren höchstens einigen Interessierten bekannt waren, wird heute damit geworben, daß vom Enkel bis zur Oma die gesamte Familie ‚drin‘ ist – im weltweiten Datennetz. Während gesellschaftliche und wissenschaftliche Auswirkungen dieser Änderungen noch nicht mal im Ansatz aufgearbeitet worden sind, hat auf technischer Seite bereits die zweite ‚Internet-Revolution‘ begonnen: die der Umwandlung des heutigen darstellungsorientierten WWW in ein Netzwerk von Bedeutungen, in ein SEMANTIC WEB. Eines bleibt jedoch unverändert: Das WWW wird nach wie vor aus verknüpften elektronischen Texten bestehen – aus sogenannten HYPERTEXTEN.

Den sich immer schneller entwickelnden Technologien muß fachwissenschaftliche Grundlagenforschung entgegen bzw. zur Seite gestellt werden. Die vorliegende Arbeit ist ein technisch orientierter Beitrag zur textlinguistischen Grundlagenforschung. Ihr Ziel ist zu zeigen, daß mit aktuellen textlinguistischen Möglichkeiten durchaus fruchtbare Untersuchungen von Hypertexten möglich sind.

1.1 Textlinguistische Motivation

Eine Frage drängt sich freilich gleich zu Beginn auf: Ist die Textlinguistik denn überhaupt für Hypertexte zuständig? Hierauf sei mit einem Bonmot aus der textlinguistischen Hypertext-Forschung geantwortet:

„Wenn sich nicht herausstellen sollte, daß HT [= Hypertext] im selben Umfang Texte sind wie Walfische Fische, dann gilt: Die Linguistik beschäftigt sich unter anderem mit Texten. HT sind Texte. Also beschäftigt sich die Linguistik mit HT.“

(Freisler 1994 S. 5)

Ob nun wirklich dort ‚Text drin ist, wo Text drauf steht‘, wird erst die genauere Betrachtung des Untersuchungsgegenstandes zeigen. Nichts scheint jedoch auf den ersten Blick näher zu liegen, als daß sich die für Texte zuständige linguistische Teildisziplin der TEXTLINGUSTIK dieser elektronischen Gebilde annimmt. Problematisch ist hierbei nur, daß die Textlinguistik noch immer um die Definition der Größe TEXT ringt. Zugegebenermaßen liegen hier die

Verhältnisse deutlich unklarer als etwa bei der linguistischen Größe SATZ. Allerdings arbeiten auch Syntaktiker nicht mit einer allgemein anerkannten Festlegung letzterer Größe, sondern mit unterschiedlichen, der jeweiligen Fragestellung angepaßten, theoretisch determinierten Satzdefinitionen – und dies mit beachtlichen Erfolgen. In der vorliegenden Arbeit wird ein KOMMUNIKATIV-FUNKTIONALER TEXTBEGRIFF vorgestellt, der sich für praktische Untersuchungen von Hypertexten eignet. Auf (durchaus existierende) theoretische Unzulänglichkeiten bzw. Gegegentwürfe wird hierbei stets hingewiesen.

Bis jetzt wurden hauptsächlich Untersuchungen zu einzelnen Aspekten der TEXTUALITÄT von Hypertexten (meist kontrastiv zu herkömmlichen Texten) durchgeführt. Im Vordergrund standen primär Probleme von KOHÄSION und KOHÄRENZ. Die vorliegende Arbeit möchte hingegen einen textlinguistischen Orientierungsrahmen für sich anschließende, breiter angelegte praktische textwissenschaftliche (siehe unten) Untersuchungen von Hypertexten liefern. Hypertexte werden hierfür in mehrere Ebenen unterteilt und auf verschiedene Aspekte hin untersucht. Ein solcher Beschreibungsapparat wird in dieser Arbeit als TAH – als TEXTLINGUISTISCHES ANALYSEMODELL FÜR HYPERTEXTE – eingeführt und exemplarisch auf dessen Anwendbarkeit hin geprüft.

1.2 Problemstellung

Als Untersuchungsgegenstand werden, wie gesagt, keine gedruckten Texte oder Diskurse der gesprochenen Sprache verwendet, sondern Hypertexte. Die Besonderheit dieser Texte ist neben ihrer elektronischen Realisierung eine tendenziell ‚nicht-lineare‘ Anordnung der Information in Teiltexthe, sogenannte KNOTEN, die miteinander über Verweise, sogenannte LINKS, verbunden sind. Bei der Lektüre entscheidet der Leser in der Regel relativ frei über die Reihenfolge der zu rezipierenden Knoten. *Relativ* deshalb, weil der Autor mit dem Setzen von Links natürlich die Zahl der NAVIGATIONS-MÖGLICHKEITEN¹ während der Lektüre begrenzt: Je mehr Links innerhalb eines Knotens existieren, um so mehr mögliche LESEPFADe kann der Leser bei der Lektüre einschlagen.

Eine allgemein akzeptierte Definition des Phänomens HYPERTEXT steht noch aus, was wiederum nur verständlich ist, da Hypertexte ja selbst (jedenfalls zum

¹ Dieser Vorgang der Textsequenzierung innerhalb eines Hypertextes durch den Leser wird oft mit *Navigieren in einem Hypertext* paraphrasiert.

größten Teil) aus der noch nicht exakt festgelegten linguistischen Größe TEXT bestehen. Ebenso wie bei der Diskussion des Textbegriffes werden für den Terminus HYPERTEXT unterschiedliche Ansätze der Sekundärliteratur kritisch vorgestellt. An dieser Stelle sollte jedoch klar sein, daß nicht auf alle offenen Fragen der HYPERTEXT-THEORIE eingegangen werden kann. Es wird sich jedoch zeigen, daß für die Struktur von Hypertexten eine relativ gut verwendbare und technisch saubere Arbeitsdefinition zu finden ist.¹ Auch die für Teil C dieser Arbeit relevanteste Komponente des Hypertextes, der LINK, ist zwar intuitiv gut faßbar, theoretisch jedoch nicht ganz einfach festzulegen. Aber auch hier läßt sich eine Definition erarbeiten, die ausreichend für die vorliegende Untersuchung ist.

1.3 Vorgehensweise

Der Untersuchungsgegenstand HYPERTEXT wird im Hauptteil A ausführlich eingeführt, das textlinguistische Instrumentarium in Teil B erarbeitet. Als Kernstück der Arbeit wird am Ende des Hauptteils B ein TEXTLINGUISTISCHES ANALYSEMODELL FÜR HYPERTEXTE (TAH) vorgeschlagen, das im folgenden Fallbeispiel des Hauptteils C zur Untersuchung der folgenden Problematik exemplarisch verwendet wird: Jeder Link stellt den Leser eines Hypertextes vor die Entscheidung, ob er im aktuellen Knoten weiterlesen oder aber dem Link folgend zu dessen ZIEL-KNOTEN² wechseln will. Dies kann unter Umständen zu einer kognitiven Mehrbelastung des Lesers führen und ihm die Orientierung in komplexeren Hypertexten erschweren.³ Eine Verbesserung der Situation wäre eventuell dadurch zu erreichen, daß man dem Leser, bevor dieser einen Link tatsächlich verfolgt, signalisiert, um was für eine Art von Knoten es sich beim Ziel des Links handelt; also etwa, ob dort ein *Beispiel* für den beschriebenen Sachverhalt des Ausgangs-Knotens zu finden ist, eine *Gegenposition* zu vorher Gesagtem o.ä. Auch Aussagen zum Medium des Ziel-Knotens sind denkbar, etwa ob es sich um einen textuellen Knoten oder ein Film- bzw. Tondokument handelt. Der Fachausdruck für solche Links mit Signalfunktion ist TYPISierter LINK. Typisierte Links werden im Rahmen des Fallbeispiels mit Hilfe von TAH untersucht: Dort wird gefragt, welche Verknüpfungsmöglichkeiten

¹ Dies freilich auf Grundlage des Textbegriffes, der in dieser Arbeit verwendet wird und auf dessen theoretische Unvollständigkeiten bereits hingewiesen wurde.

² Die Corpus-Hypertexte haben keine Links, die über mehr als einen Zielknoten verfügen. Mit neuen Technologien ist dies jedoch denk- und machbar, mehr dazu auch in Kapitel 15.2.3.

³ Siehe hierzu Kapitel 6.4.2.

rhetorischer, semantischer und/oder logischer Natur prinzipiell zwischen Textteilen existieren und ob, bzw. wie, diese zu klassifizieren sind. Herangezogen werden hierfür linguistische und informationswissenschaftliche Typologien sowie die proprietären Typisierungsansätze der Corpus-Hypertexte.

1.4 Einschränkungen

Vorliegende Untersuchung ist, wie gesagt, eine genuin textlinguistische. Dies scheint auf den ersten Blick insofern problematisch, als daß sich seit längerem eine tendenziell interdisziplinäre Textwissenschaft¹ auszubilden versucht, die neben der Linguistik auch Disziplinen wie *Psychologie*, *Neurobiologie*, *Soziologie* und weitere Teilbereiche integriert. Diese Herangehensweise erhält durch das sich in (bestimmten) sprachwissenschaftlichen Kreisen immer mehr durchsetzende Selbstverständnis der Linguistik als Teildisziplin einer allgemeinen Kognitionswissenschaft² vermehrt Rückendeckung. Auch der Untersuchungsgegenstand HYPERTEXT ist natürlich für die unterschiedlichsten Disziplinen relevant, wie die Veröffentlichungsflut aus BWL, Soziologie, Philosophie, Literaturtheorie u.ä. zu diesem Thema zeigt (siehe Kapitel 3).

Meiner Ansicht nach hängt ein interdisziplinäres Projekt bei allen möglichen Synergieeffekten jedoch stark vom Input der Einzeldisziplinen ab. Die Textlinguistik als relativ neue sprachwissenschaftliche Teildisziplin kann im Moment noch nicht die Ergebnisse liefern, die linguistische Grundlage für eine fächerübergreifende Untersuchung von kognitiver TEXTVERARBEITUNG oder gar HYPER-TEXT-VERARBEITUNG sein könnte. Auf die Integration des deskriptiven Analysemodells in ein umfassendes kognitives Modell wurde daher bewußt verzichtet, TAH soll als rein textlinguistischer Beitrag zur Beschreibung hyper-textueller Funktionen und Strukturen gesehen werden.³ Insofern versucht die vorliegende Untersuchung auch, einen (bescheidenen) Teil der noch ausstehenden ‚Hausaufgaben‘ aus dem Bereich der Textlinguistik zu erledigen. Die Ergebnisse könnten dann ihrerseits Input für empirische Untersuchungen und angewandte Forschung sein, etwa für die Implementierung einer semantisch angereicherten Hypertext-Oberfläche (siehe Kapitel 15.1.3) unter Benutzung

¹ Vgl. hierzu etwa Kapitel 1 von van Dijk 1980a, Strohner 1990 sowie Friederici 1999.

² Vgl. hierzu etwa Rickheit u. Strohner 1993, Sucharowski 1996 sowie Pörings u. Schmitz 1999.

³ Auf eine mögliche, voraussichtlich unproblematische Einbettung des deskriptiven TAH in ein größeres kognitives Erklärungsmodell der Hypertext-Verarbeitung wird im Kapitel *Ausblick* dieser Arbeit hingewiesen.

neuer Texttechnologien wie XLINK, XML-TOPIC-MAP und weiterer XML-DERIVATE.

1.5 Zielgruppe

Die Arbeit wendet sich explizit an ein linguistisches Fachpublikum mit gewissen interdisziplinären Neigungen sowie einem besonderen Interesse an moderner Texttechnologie. Allgemeine linguistische Fachbegriffe werden daher ohne gesonderte Vorststellung verwendet. Textlinguistische Termini werden hingegen ausführlich eingeführt und kritisch besprochen. Für das Verständnis ist eine gewisse Vertrautheit mit dem Internet sowie die Kenntnis informationswissenschaftlicher Grundbegriffe hilfreich – aber nicht unabdingbar. Die für die Arbeit relevanten technischen Konzepte, Implementierungen und Spezifikationen werden von Grund auf eingeführt. Hierbei wird, soweit dies die Komplexität der Materie ermöglicht, ‚behutsam‘ vorgegangen – ohne jedoch übermäßig stark zu vereinfachen. Der technisch interessierte Leser wird hauptsächlich in den Fußnoten auf Einschränkungen, Problembereiche und weiterführende Literatur aus der Informatik verwiesen. Die wichtigsten technischen Begriffe sind darüber hinaus in einem Glossar aufgeführt.

Die Arbeit dürfte durchaus auch für Informatiker interessant sein, die gewillt sind, sich mit sprachwissenschaftlichen Fragestellungen auseinanderzusetzen. In terminologischen Nöten sei auf die allgemeinen linguistischen Nachschlagewerke aus Kapitel 3.2 verwiesen.

Leser, die mit den Konzepten von Hypertext und den Implementierungen von Markup-Sprachen bereits vertraut sind, können die ausführlichen informationswissenschaftlichen Einführungen des Kapitels 15 getrost ‚querlesen‘.

2 Angaben zum Corpus

Nachdem die allgemeine Ausrichtung der Untersuchung im letzten Kapitel erläutert wurde, sollen nun die Präferenzen genannt werden, die für die Auswahl der Hypertexte des Corpus' von Bedeutung waren.

2.1 WWW als technische Plattform

Bei der Auswahl des Corpus' müssen gewisse Festlegungen gemacht werden, um die Vergleichbarkeit der Corpus-Hypertexte zu gewährleisten: Zuerst sollte man sich auf eine mediale Realisierung, eine technische Plattform der zu untersuchenden Hypertexte einigen. Zahlreiche Softwaresysteme stehen nämlich zur Implementierung eines Hypertextes zur Verfügung (vgl. Schnupp 1992 S. 97 f.). Als derzeit mit Abstand bedeutendste elektronische Realisierung darf das WORLD WIDE WEB (WWW) gelten.¹ Alle anderen Hypertext-Systeme führen heute mehr oder weniger ein Nischendasein.² Aus diesem Grund wird sich die Arbeit ausschließlich mit im WWW präsenten und in der Sprache HTML³ geschriebenen Hypertexten beschäftigen. Nur bei relevanten Besonderheiten anderer Systeme wird auf diese Implementierungen kurz eingegangen.

2.2 Informationstexte als primäre Textsorte

Die zweite Einschränkung betrifft die TEXTSORTEN⁴, die es zu untersuchen gilt. Im WWW sind verschiedene Textsorten zu finden, HYPERTEXT selbst hingegen ist nicht als eine Textsorte zu interpretieren (siehe Kapitel 10.5)! Zweifelsohne werfen die verschiedenen, hypertextuell realisierten Textsorten jeweils spezifische textlinguistische Fragestellungen auf. Zu denken ist hier auch an die unter dem Begriff HYPERFICTION bekanntgewordene, elektronisch realisierte Belletristik, die versucht, im neuen Medium einen ästhetischen Mehrwert zu erlangen.

¹ Die Entwicklung des Word Wide Web und des Internet wird in Kapitel 4.2 zusammenfassend beschrieben.

² Drei Gründe lassen sich nennen, wieso Hypertext-Realisierungen überwiegend im WWW stattfinden: Die Hypertexte können mit der im Hinblick auf das Betriebssystem unabhängigen Hypertext-Beschreibungssprache HTML auf einer nahezu beliebigen Rechnerarchitektur implementiert werden. Die erstellten Hypertexte werden in das weltweite Computernetz INTERNET eingespielt, das noch immer mit einem exponentiellen Wachstum der Benutzerzahlen aufwarten kann. Die zur Ansicht von WWW-Hypertexten notwendige Browser-Software wird von den meisten Firmen aus Marketinggründen verschenkt (vgl. auch: Münz 1997a [htxt212.htm]).

³ Eine kurze Einführung in HTML gibt das Kapitel 15.2.

⁴ Sämtliche textlinguistischen Begriffe werden ausführlich in Teil B eingeführt.

Für die vorliegende Untersuchung werden jedoch hauptsächlich Informations- und Instruktionstexte herangezogen, die sich mit Informationstechnologien oder linguistischen Fragestellungen beschäftigen. Diese Einschränkung hat mehrere Ursachen methodischer, praktischer und persönlicher Natur: Zum einen ist die Spezialisierung auf diese Textsorte durch den beruflichen Werdegang des Autors zu erklären, der sich über mehrere Jahre hinweg mit dem Verfassen von Anleitungen und Lernmedien für Software sowie mit Programmierung beschäftigt hat. Schon aus diesem Grund lag es nahe, vorhandenes Expertenwissen für eine spezielle Textsorte in eine textlinguistische Untersuchung einfließen zu lassen. Andererseits wurden gerade Informations- und Instruktionstexte von Anfang an massiv als Hypertexte realisiert und machen auch heute (noch) die Mehrheit der Hypertexte im WWW aus.¹ Auch lautet eine Hypothese des Autors, daß solche Textsorten tendenziell geeigneter für eine semanto-pragmatische Typisierung der Links sind, da sie vermutlich vermehrt über standardisierte inhaltliche Verknüpfungen zwischen Textteilen (bzw. Knoten) verfügen. Ein Beispiel aus dem Hypertext SELF des Corpus' kann dies veranschaulichen:²

„Spaltenabstand (column-gap)
Anzeigebeispiel: So sieht's aus

Sie können bestimmen, wie groß der Abstand zwischen den einzelnen Spalten sein soll. Bei mehr als zwei Spalten gilt diese Angabe für alle Spaltenabstände.

Beispiel (Style-Sheet-Definition für HTML-Tag im Text):
<div style="columns:3; column-gap:0.5cm">Inhalt</div>

Erläuterung:

Mit column-gap: können Sie den Spaltenabstand bestimmen. Erlaubt ist eine numerische Angabe.“

(SELF [tdci.htm] Abschnitt: *Spaltenabstand (column-gap)*)

Die obige Erklärung gliedert sich in mehrere Teile:

- Eine Einleitung, die anreißt, um was es überhaupt geht (*Sie können bestimmen, wie ...*)
- Ein Beispiel, welches einen exemplarischen Quellcode zeigt, sowie ein Anzeigebeispiel, welches die Auswirkungen des eben genannten Quellcodes

¹ In Kapitel 4.2 wird über die Motivationen zur Gründung des WWW zu lesen sein, daß diese primär darin bestand, das „problem of information access“ zu lösen (vgl. Berners-Lee 1989).

² Die Unterstreichungen symbolisieren Links (siehe hierzu auch den Abschnitt *Typographische Konventionen* dieser Arbeit).

auf dem Bildschirm illustriert. Das Anzeigebeispiel ist in einen separaten Knoten ausgelagert und kann per Klick auf den LINK *So sieht's aus* aufgerufen werden.

- Eine abschließende Erläuterung mit weiterführenden Informationen.

Schon nach einem kurzen Blick auf lediglich ein Beispiel kann vermutet werden, daß Teile von Informationstexten aus der Technischen Dokumentation wiederum aus bestimmten Teiltextrn bestehen, die sich einer Typisierung nicht von vornherein verschließen dürften. Im obigen Fall könnte man etwa die ad hoc gebildeten LINK-Etiketten *Visualisierung* (im obigen Beispiel: *So sieht's aus*) oder *Zusatzinfo* (im obigen Beispiel: *numerische Angabe*) einführen. Ob diese ersten Vermutungen Bestand haben, wird sich jedoch erst nach umfangreicheren theoretischen und praktischen Untersuchungen feststellen lassen.

2.3 Die Texte des Corpus'

Wichtigster Hypertext des Corpus' ist zweifelsohne der Titel *SelfHTML. Die Energie des Verstehens: HTML-Dateien selbst erstellen* von Stefan Münz, der unter dem Kürzel SELF verwendet wird. Besonders geeignet ist dieser Text auch aufgrund seiner hohen Link-Dichte im Fließtext. SELF wurde im Jahr 1998 unter der Adresse <http://www.teamone.de/selfhtml/> online gestellt. Kurz vor Abgabe der vorliegenden Untersuchung ist jedoch eine Neuauflage unter der Adresse <http://selfhtml.teamone.de/> ins WWW gestellt worden. Die alte Adresse ist nun nicht mehr gültig. Durch die Neuauflage bot sich die Chance, die beiden Versionen vergleichend zu untersuchen. Dies ist besonders interessant, da sich die in SELF benutzten typisierten LINKS zum Teil geändert haben. Aus diesen Gründen ist sowohl die alte als auch die neue Auflage in das Corpus aufgenommen. Die Neuauflage wird im Rahmen der vorliegenden Arbeit mit SELF8 abgekürzt, da diese mit der Versionsnummer 8.0 versehen wurde. Die alte Version befindet sich seit kurzem wieder auf dem neuen Server des SelfHTML-Teams, und zwar im Archiv unter <http://selfaktuell.teamone.de/archiv/doku/7.0/>. So bleibt die Überprüfbarkeit der Analyse der alten Version gewährleistet.

Ein weiterer Text kommt aus der wissenschaftlichen Fachkommunikation: Das Grammatik-Informationssystem GRAMMIS, welches unter diesem Kürzel verwendet wird. GRAMMIS ist insofern interessant, als daß dieses Projekt in zweifacher Weise TYPISIERTE LINKS verwendet: Zum einen sind die LINKS durch

farbige Markierungen bestimmten Bereichen zugeordnet. Zum anderen wird diese Zuordnung zum Teil durch MOUSE-OVER-TEXTE¹ unterstützt.

Ein Hypertext, der thematisch aus dem Rahmen fällt, ist das Sportportal der Sport 1 GmbH, das unter dem Kürzel SPORT1 verwendet wird. Dieser Hypertext lohnt aufgrund seiner zahlreichen PIKTOGRAMME², die für die unterschiedlichen LINK-Typen verwendet werden, einer genaueren Betrachtung. Außerdem stellt SPORT1 eine im WWW im häufiger anzutreffende Sorte von Hypertexten dar, die durch eine geschickte Verquickung von Informations- und Werbetexten gekennzeichnet ist. Hier können zum Teil sehr interessante Unterschiede bzgl. Funktion und Struktur einzelner Knoten im Vergleich zu den restlichen, rein informativen Hypertexten des Corpus' herausgearbeitet werden.

Ebenfalls als Primärtext dient die Informationsseite des W3C. Der besondere Reiz dieses Hypertextes resultiert nicht nur aus der Tatsache, daß das W3C die Standards im Internet festlegt und verwaltet – und damit als die maßgebende Informationsseite zur Internet-Technologie gesehen werden muß –, sondern auch aus der Herausforderung für die W3C-Verantwortlichen, im Rahmen der sich schnell ändernden Technologien, unterschiedliche Versionen von technischen Spezifikationen – und damit von Hypertext-Knoten – zu verwalten und, im Idealfall, diese zeitlichen Relationen dem Leser auch zu visualisieren.

Folgende Tabelle stellt die Corpus-Hypertexte nochmals übersichtlich zusammen:

Kürzel	Name	Adresse
GRAMMIS	Grammis. Das grammatische Informationssystem des IDS	http://hypermedia.ids-mannheim.de/grammis/
SELF	SelfHTML. Die Energie des Verstehens: HTML-Dateien selbst erstellen (Version 7)	http://selfaktuell.teamone.de/archiv/
SELF8	SelfHTML. Die Energie des Verstehens: HTML-Dateien selbst erstellen (Version 8)	http://selfhtml.teamone.de/
SPORT1	Sport 1 (Sportportal)	http://www.sport1.de/
W3C	W3C. World Wide Web Consortium	http://www.w3c.org/

Tabelle 1: Hypertexte des Corpus'

¹ Darunter versteht man Texte, die auf dem Bildschirm erscheinen, sobald der Mauszeiger auf einen Link bewegt wird.

² PIKTOGRAMME sind kleine, ikonenhafte Bilder, die zur Visualisierung von Link-Typen benutzt werden können. Sie werden in Kapitel 15.2.4 vorgestellt.