

# Ein wortnetzbasierter Ansatz für die korpusgestützte Modellierung von Fachterminologie

## Ein Beitrag zur digitalen Fachlexikographie

**Michael Beißwenger (Dortmund)**

---

**Preprint.** Erscheint in: In: *Zeitschrift für germanistische Linguistik* 38 (2010), S. 346-369.

---

1. Einleitung
2. Terminologische Diversität als Herausforderung für die Modellierung von Fachwortschätzen
3. Der TermNet-Ansatz
  - 3.1 Projektkontext und Anwendungsszenario
  - 3.2 Modellierung des terminologischen Wortnetzes
  - 3.3 Zwei Anwendungsbeispiele
4. Fazit
5. Literatur

### **Abstract**

This article describes an approach for modelling domain-specific terminology in a wordnet-style representation. It uses the fundamental entities and relations introduced for the Princeton WordNet (Fellbaum 1998) and expands upon these in a way that fits for the representation of technical terms that are given in a corpus with scientific texts.

The article starts with an overview on some essential semantic and lexical features of technical terms and terminological systems from the perspective of LSP research and formulates a set of requirements that derive from these characteristics for a modelling of domain-specific terminology which also aims to include terminological diversity (i.e. the existence of several terminological systems competing within the same special-field domain).

Subsequently, the modelling approach with its fundamental modelling units will be introduced and the essential modelling decisions made with regards to the previously formulated requirements will be illuminated.

In closing, two applications based on the modelling approach will be introduced: on the one hand a hypertext glossary of the domains "hypermedia research and text technology" which has been built in the context of the DFG-funded project "Text-grammatical foundations of text-to-hypertext conversion" at TU Dortmund University, and on the other hand the component "Grammatische Ontologie" of the grammatical online information system "Grammis" at the Institute for German Language (IDS), Mannheim.

## **1. Einleitung**

Dieser Beitrag stellt einen Ansatz für die Modellierung von Fachterminologie vor, der darauf zielt, die in einer Fachdomäne verwendete Terminologie so zu beschreiben, dass daraus Wissen über die sprachliche Strukturierung ihrer fachlichen Gegenstände in dreierlei Hinsicht erschlossen werden kann:

- (i) Wissen über die Verwendung einzelner Termini in Fachtexten der Domäne;

- (ii) Wissen über die Beziehungen einzelner terminologischer Konzepte zu ihren Nachbarkonzepten;
- (iii) Wissen um die Existenz konkurrierender terminologisch-konzeptueller Systeme zu bestimmten Ausschnitten aus dem fachlichen Objektbereich der Domäne und die Bezüge zwischen ihnen.

Der Ansatz ist

- fachlexikographisch motiviert, insofern er darauf zielt, Terminologiesysteme und terminologische Diversität in einer Weise zu beschreiben, dass aus den entsprechenden Ressourcen digitale Glossare generiert werden können, die sowohl textbezogen (zur Bewältigung terminologiebedingter Störungen bei der Textrezeption) als auch lernbezogen (zur terminologiegeleiteten Einarbeitung in die konzeptuelle Strukturierung des fachlichen Objektbereichs der Domäne) genutzt werden können;
- wortnetzbasierter, insofern er auf die Modellierungsprinzipien von Wortnetzen im Stile des Princeton WordNet (Fellbaum 1998) aufbaut und diese terminologiespezifisch erweitert;
- korpusgestützt, insofern die im terminologischen Wortnetz beschriebenen Termini (intellektuell) auf Basis eines digitalen Fachtextkorpus ermittelt und die Einträge im Wortnetz (semi-automatisch) mit Textstellen im Fachtextkorpus vernetzt werden, in denen die betreffenden Termini definiert und verwendet sind.

Der Ansatz wurde entwickelt im Rahmen des von Angelika Storrer 2002–2003 am IDS Mannheim und 2003–2009 an der TU Dortmund geleiteten Teilprojekts „Hypertextualisierung auf textgrammatischer Grundlage“ (HyTex, <http://www.hytex.info>) der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“ (<http://www.text-technology.de>). Gegenstand des Projekts war die Entwicklung von Strategien und Methoden zur semi-automatischen Generierung von Hypertext-Sichten auf ursprünglich linear organisierte Fachtexte auf der Grundlage von in die Dokumente eingebrachten linguistischen Annotationen und einem terminologischen Wortnetz zur in der Domäne verwendeten Fachterminologie (Storrer 2010). Als Anwendungsszenario diente die Idee einer Hypertext-Leseumgebung für Fachtexte, die über semi-automatisch generierte Link- und Navigationsangebote die Kohärenzbildung eines selektiv und quer lesenden Nutzers unterstützt. Fokussiert wurde dabei auf solche Nutzer, die sich unter Zeitdruck zielgerichtet Wissen zu einer Domäne erarbeiten müssen, in der sie nicht Experten sind, sondern Semi-Experten (z.B. Experten angrenzender Disziplinen, Wissenschaftsjournalisten, Fachübersetzer; vgl. Bergenholtz & Kaufmann 1997: 99-102) oder Novizen (Studierende derjenigen Fächer, die eine bestimmte Fachdomäne an einer Hochschule institutionell vertreten). Diese rezipieren Fachtexte nicht nur häufig selektiv, sondern verfügen darüber hinaus i.d.R. über eine nur partielle Orientierung über die in der betreffenden Domäne verwendete Terminologie. Sie sind daher in besonderem Maße vom Explizithedilemma fachsprachlicher Texte betroffen: Wo Texte nur in Teilen rezipiert werden können, muss fachliche Kommunikation, um Verstehen sicherzustellen, in besonderem Maße explizit sein; umgekehrt wird gerade „Explizithed als gelegentlich erstrebte Eigenschaft von Fachtexten [...] auf der anderen Seite bezahlt mit Länge, Umständlichkeit und Redundanz des Textes“ (von Hahn 1997: 383): Texte werden unhandlich, wollte der Autor an jeder Stelle seiner Verwendung die für die Seman-

tisierung eines Terminus relevante Definition wiederholen. Nach Möhn (1997: 155) ist die

Ökonomie der Expertenkommunikation [...] vor allem dadurch gegeben, daß die Kompetenz des fachspezifischen Referierens vorausgesetzt werden kann. Ein redundantes Paraphrasieren wird als Verstoß gegen die Gruppenkonvention gewertet.

Die Funktion von Termini besteht in diesem Zusammenhang gerade darin, einen – u.U. theoretisch abstrakten – fachlichen Wissensausschnitt in einer bestimmten Konzeptualisierung kompakt zu repräsentieren und auf diese Weise die Kommunikation und Reflexion über theoretisch komplexe Sachverhalte zu vereinfachen. Das Verständnis eines Terminus setzt dabei häufig voraus, nicht nur die zugehörige Definition zu kennen, sondern auch über dessen Stellung zu seinen terminologischen Nachbarkonzepten und die Struktur des zugehörigen terminologischen Netzes (des betreffenden Autors oder wissenschaftlichen Paradigmas) orientiert zu sein.

Entsprechend stehen gerade Semi-Experten und Novizen oftmals vor dem Problem, mit Texten umgehen zu müssen, die für einen Adressatenkreis verfasst wurden, dem der Autor ein deutlich höheres Kenntnisniveau zuspricht (und in denen er terminologiebezogene Wissensvoraussetzungen nicht an jeder Stelle, an der diese verstehensrelevant sind, textuell expliziert). Dadurch verstärkt sich das Problem fehlender oder nur unzureichender terminologischer Orientierung bei der Fachtextrezeption. Terminologische Komplexität und Diversität im wissenschaftlichen Diskurs einer Domäne stellt dabei zugleich ein Kommunikations- wie auch ein Informationsproblem (i.S.v. Fluck 1996: 37f.) dar: Unzureichende Orientierung über die Bedeutung von Termini und ihre Stellung im Fachwortschatz der Domäne führt zu Verstehensproblemen bei der Textrezeption und dadurch zu Problemen beim Aufbau individueller Wissensstrukturen zu den wissenschaftlichen Gegenständen und ihrer reflexiven Durchdringung.

Abschnitt 2 gibt unter fachsprachenlinguistischer Perspektive einen Überblick zu einigen zentralen lexikalisch-semantischen Besonderheiten von Fachwortschätzen und motiviert die Anforderungen an eine Modellierung von Fachterminologie, die insbesondere auch terminologische Diversität umfassen soll.

Abschnitt 3 skizziert den Modellierungsansatz, der im HyTex-Projekt gewählt wurde und erläutert mit Blick auf die in Abschnitt 2 beschriebenen Anforderungen die wichtigsten dabei getroffenen Modellierungsentscheidungen.

Abschließend werden zwei Anwendungen vorgestellt, denen der Modellierungsansatz zugrunde liegt: ein Hypertext-Glossar zu den Domänen „Hypermediaforschung und Texttechnologie“, das im Rahmen des HyTex-Projekts in einem Demonstrations-Prototypen implementiert wurde, sowie die Komponente „Grammatische Ontologie“ des Grammatischen Informationssystems *Grammis* des Instituts für Deutsche Sprache (IDS), Mannheim.

## **2. Terminologische Diversität als Herausforderung für die Modellierung von Fachwortschätzen**

Fachwortschätze weisen gegenüber gemeinsprachlichen Wortschatzbereichen eine Reihe semantischer und lexikalischer Besonderheiten auf, aus denen sich spezifische Anforderungen an eine deskriptive Modellierung von Fachterminologie ableiten lassen:

- (i) Fachtermini sind definitiv fixiert und sind Handlungsmittel zur Konstruktion wissenschaftlicher Gegenstände.

Die Angabe einer Definition zu einem Terminus im Fachtext dient – sofern der Autor kenntlich macht, dass er diese Definition teilt – der expliziten Festlegung einer Wortbedeutung und damit der Verstehenssicherung. Ein Autor, der eine Definition zu einem sprachlichen Ausdruck gibt, legt sich, soweit nicht anders expliziert, für sein Folgehandeln auf eine bestimmte Verwendung des betreffenden terminologischen Ausdrucks fest. Ein Leser, der seinen Textäußerungen in der von ihm intendierten Abfolge folgt und über das von ihm vorausgesetzte Vorwissen verfügt, wird im Falle erneuter Verwendung des terminologischen Ausdrucks darüber orientiert sein, wie dieser Ausdruck zu semantisieren ist.

Die genuine Prägung eines Terminus (d.h.: die explizite und erstmalige definitivische Verbindung eines bestimmten sprachlichen Ausdrucks mit der Beschreibung eines bestimmten Konzepts) zielt darauf, einen spezifisch neuen Zugang zur Strukturierung eines wissenschaftlichen Gegenstands zu eröffnen, der ohne die durch ihn repräsentierte charakteristische Verbindung von Form und Konzept im wissenschaftlichen Diskurs der Domäne noch nicht als Erkenntnis- und Kommunikationsmittel vorhanden ist. Mit Wiegand (1996: 72) können Akte dieser Art als *systembezogene Nominationshandlungen* (oder: *systembezogene Erstbenennungshandlungen*) bezeichnet werden, also Handlungen, die mit dem Ziel ausgeführt werden,

eine bis dahin – nach dem Wissensstand des Handlungsausführenden – noch gar nicht benannte Entität, deren ontologischer Status gleichgültig ist, dadurch einer gleichzeitig mitbenannten Klasse zuzuordnen, daß man der noch nicht benannten Entität mittels einer dazu geeigneten Äußerung einen Klassennamen (oder: einen Gattungsnamen i.S.v. *nomen appellativum*) zuweist, den sie nach der Ausführung der Nominationshandlung möglichst dauerhaft und nicht nur für die kurze Dauer derjenigen Interaktionseinheit tragen soll, zu der die ausgeführte systembezogene Nominationshandlung gehört.

Obwohl „Wissenschaftler in einem relativ weitgespannten Rahmen Herr des terminologischen Lexikons“ sind (ebd.: 98) und somit die von ihnen verwendeten wissenschaftlichen Termini grundsätzlich jederzeit nach Belieben neu einführen können, geschieht Terminologieprägung in aller Regel nicht unmotiviert und unkontrolliert, sondern häufig entweder – sprachreflexiv – im Kontext einer (kritischen) Diskussion der in der Domäne bereits vorhandenen Terminologie oder aber – konzept- bzw. gegenstandsbezogen – im Kontext einer kritischen Diskussion des aktuellen Stands des wissenschaftlichen Erkenntnisprozesses oder der wissenschaftlichen Zugänge zum entsprechenden fachlichen Gegenstand. Für ein tiefergehendes Verständnis eines Fachterminus (der – im Sinne einer instrumentalistischen Sprachauffassung – stets Handlungsmittel und nicht lediglich Etikett für eine a priori gegebene Wirklichkeit ist) ist es daher oftmals gerade wichtig, nicht nur die vom Urheber formulierte Definition zu kennen, sondern auch die wissenschaftliche Argumentation und die Motivation nachzuvollziehen, auf Basis derer sich der Urheber dazu entschlossen hat, die betreffende Sache auf eine bestimmte (im Terminus fixierte) Weise zu konzeptualisieren und dieses Konzept mit einem bestimmten Namen zu versehen.

Anforderung an die Modellierung: Eine Terminologiemodellierung, die die spezifische Verwendung von Termini im wissenschaftlichen Diskurs der Domäne rekonstruierbar machen möchte, muss die Repräsentation terminologischer Einträge im Modell rückbinden an die einschlägigen Textstellen in den entsprechenden Fachtexten.

- (ii) Fachtermini sind häufig eingebunden in terminologische Systeme und gewinnen ihre spezifische Funktion durch Opposition zu den übrigen Systemeinheiten.

Die Prägung neuer wissenschaftlicher Termini geht i.d.R. einher mit der Motivation, einen Ausschnitt aus dem Objektbereich zu systematisieren: Häufig prägt ein Autor nicht nur einen neuen Terminus, sondern führt in ein- und demselben Kontext (z.B. dem entsprechenden Abschnitt eines wissenschaftlichen Fachartikels) gleich mehrere neue Termini ein, die insgesamt eine konzeptuelle Gliederung des betreffenden Objektbereichsausschnitts entwerfen (z.B. bei der Vorstellung eines Klassifikations- oder Typologisierungsvorschlags für denselben) oder aber ein im Fokus der Betrachtung stehendes Konzept durch eine Charakterisierung der Nachbarkonzepte, von denen es abzugrenzen ist, für die Adressaten zu schärfen.

Beispiel 1 ist einer metalexikographischen Arbeit entnommen, die dem Konzept der lexikographischen Mikrostruktur gewidmet ist:

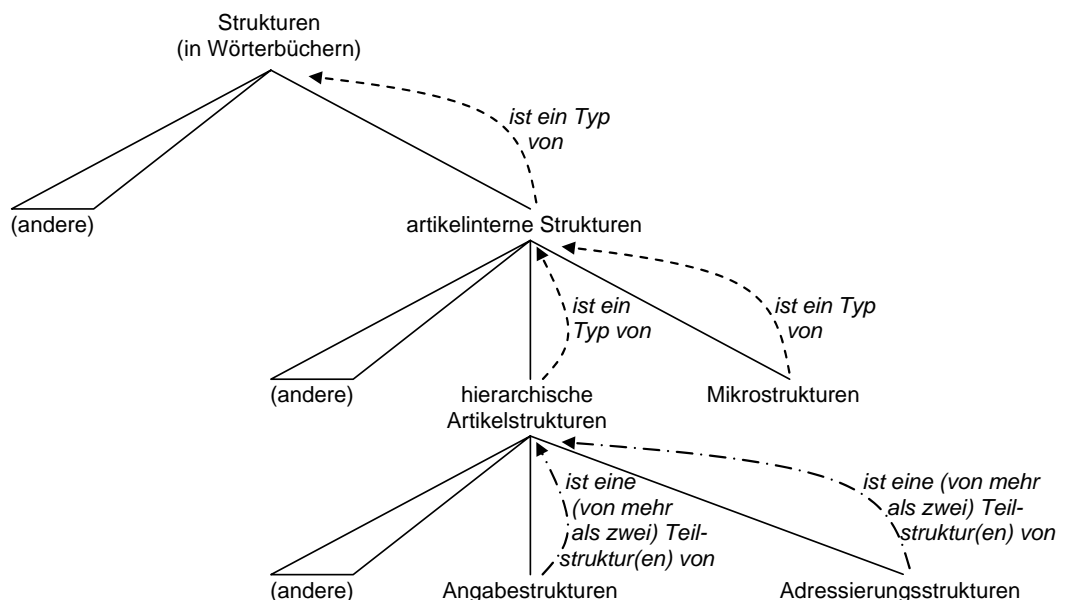
**Beispiel 1:** Textausschnitt aus Wiegand (1989: 440); Hervorhebungen MB:

Für eine differenzierte und explizite metalexikographische Textdeskription reicht es nicht aus, wenn man nur über den Begriff der (bzw. einen Begriff von) Mikrostruktur verfügt. Es sind weitergehende Unterscheidungen erforderlich, die dazu führen, daß Mikrostrukturen von standardisierten Wörterbuchartikeln als Strukturen unter anderen zu gelten haben [...].

Nachfolgend werden daher Mikrostrukturen nicht mehr isoliert betrachtet, sondern „im Zusammenhang“ mit anderen artikelinternen Strukturen wie der hierarchischen Artikelstruktur und zwei ihrer Teilstrukturen, der Angabestruktur und der Adressierungsstruktur.

In Anlehnung an das Wortfeldkonzept und den Saussure'schen *valeur*-Begriff kann man mit Blick auf Operationen wie diejenigen in Beispiel 1 von einer textuellen Etablierung eines terminologischen Feldes sprechen, die beansprucht, bestimmte empirisch zugängliche Phänomene (hier: Strukturen in gedruckten Wörterbuchartikeln) (i) als wissenschaftlichen Gegenstandsbereich zu konstituieren und diesen durch Einführung einer Reihe von sprachlichen Etiketten für einzelne Phänomentypen kategorial zu strukturieren, und zwar so, dass die einzelnen Einheiten ihre charakteristische Stellung im System durch ihre spezifische Stellung zu (Abgrenzung von) den anderen im System vorgesehenen Einheiten gewinnen. Die Konzept-Etiketten *artikelinterne Strukturen*, *hierarchische Artikelstrukturen*, *Mikrostrukturen*, *Angabestrukturen* und *Adressierungsstrukturen* können in diesem Sinne als systematisch voneinander abgegrenzte Einheiten zur Strukturierung eines im Textausschnitt nicht mit allen seinen Elementen eingeführten terminologischen Feldes „Strukturen in Wörterbüchern“ aufgefasst werden. Ohne dass im Textausschnitt bereits präzise definitorische Konzeptualisierungen zu den einzelnen Phänomentypen gegeben werden (dies wird an späterer Stelle im betreffenden Artikel geleistet), werden die mit ihren terminologischen

Bezeichnungen eingeführten Konzepte in eine hierarchische Ordnung gebracht, um dem Rezipienten (bei dem eine Vertrautheit mit diesen Konzepten noch nicht vorausgesetzt werden kann) die Möglichkeit zu eröffnen, den Bereich in seinem mentalen (Fach-)Lexikon konzeptuell zu verorten und für dessen genauere Strukturierung Unterknoten anzulegen, deren Eigenschaften zwar noch nicht en detail bekannt sind, die aber bereits so umrissen sind, dass die Art und Weise ihrer hierarchischen Unterordnung unter den bereits vorhandenen Knoten klar ist: artikelinterne Strukturen (und weitere, noch nicht näher bezeichnete Strukturen) sind Typen von Strukturen in Wörterbüchern, hierarchische Artikelstrukturen und Mikrostrukturen (und weitere, noch nicht näher bezeichnete Strukturen) sind Typen von artikelinternen Strukturen, und Angabestrukturen und Adressierungsstrukturen (und weitere, noch nicht näher bezeichnete Strukturen) sind Teilstrukturen von hierarchischen Artikelstrukturen (vgl. die Veranschaulichung in Abb. 1).



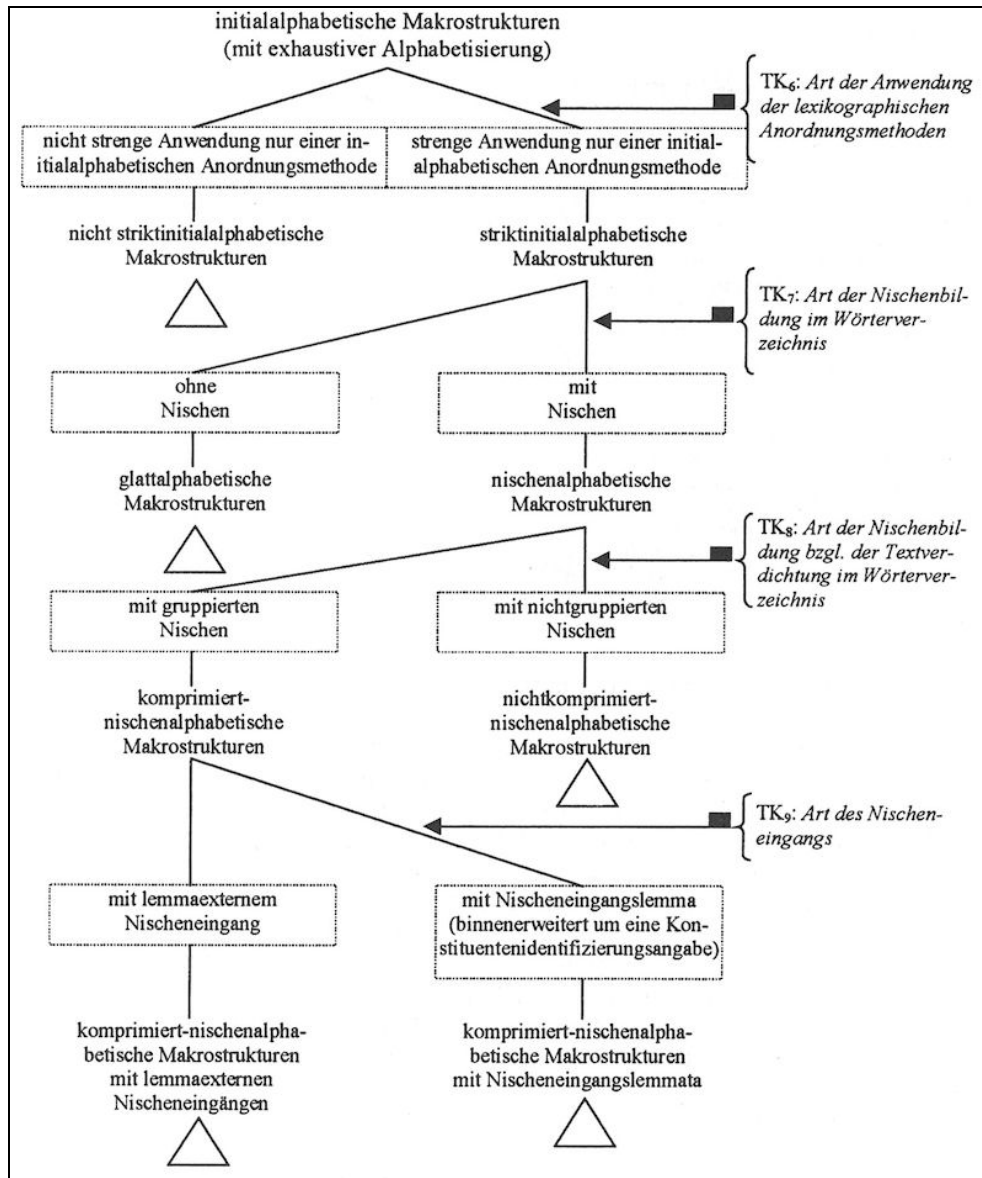
**Abb. 1:** Ausschnitt aus dem terminologischen Feld „Strukturen in Wörterbüchern“, der aus dem Textausschnitt in Beispiel 1 erschlossen werden kann.

In Bezug auf die Gewinnung wissenschaftlichen Wissens und seine Vermittlung inner- und außerhalb der Domäne haben terminologische Systeme nach Budin (1996) Ordnungsfunktion in dreierlei Hinsicht: (1) Sie dienen als Instrumente, fachliches Wissen kategorial zu organisieren und neue wissenschaftliche Erkenntnis zu generieren (epistemische Funktion); (2) sie sind „Ordnungsmittel für die praktische Informationsvermittlung in Form von Thesauri, Klassifikationssystemen etc.“ (informationelle Funktion); (3) sie tragen dazu bei, fachliche Kommunikationsprozesse begrifflich zu strukturieren (diskursleitende Funktion).

Für die Erfüllung aller drei Funktionen ist – insbesondere im Falle von Termini, die Knoten in hierarchisch strukturierten Klassifikations- oder Typologiesystemen bezeichnen – eine motivierte Konstruktion der zu ihrer Bezeichnung vorgesehenen terminologischen Ausdrücke hilfreich. Mehrgliedrige Komposita und um

mehrere Attribute ausgebaute Nominalphrasen sind aufgrund der Wortbildungsmöglichkeiten des Deutschen in deutschen Wissenschaftssprachen hoch frequent (Fluck 1996: 50ff.; Fraas 1997: 429; Roelcke 2005: 73f.), um die zur Gewinnung der Knoten in mehrstufigen Klassifikations- und Typologiesystemen angewandten Klassifikations- bzw. Typologiekriterien im terminologischen Ausdruck selbst abzubilden. Eine „Selbstdeutigkeit von Komposita [...] oder Mehrwortlexemen“ im Sinne einer „Erschließbarkeit der Bedeutung eines Terminus und einer Stelle in einem Begriffssystem aus seiner Oberflächenform“ (von Hahn 1997: 385f.) kann hierbei freilich nicht vollständig erreicht werden (schon alleine aufgrund der Nicht-Kompositionalität der Relationen zwischen Erst- und Zweitglied in Determinativkomposita). Dennoch hat eine „Aufnahme der Bezeichner der wichtigsten Begriffs-„Stationen“ in den komplexen Terminus“ allein schon unter mnemotechnischer Perspektive als günstig zu gelten (ebd.: 386) und wird entsprechend häufig angewandt, wie etwa in der Typologiedarstellung aus einer metalexikographischen Arbeit in Beispiel 2, mit welcher eine Reihe von Strukturtypen eingeführt und analog zu den je Typologisierungsschritt zusätzlich angewandten Typologiekriterien mit einem gegenüber dem jeweils übergeordneten Knoten erweiterten Ausdruck bezeichnet werden.

**Beispiel 2:** „Zweiter Ausschnitt aus einer Typologie von Makrostrukturen von gedruckten Nachschlagewerken“ (Abb. 2-8 aus Wiegand 2002: 432):



Anforderung an die Modellierung: Termini sollten in ihren spezifischen Beziehungen zu anderen Einheiten des Terminologiesystems dargestellt werden, dem sie angehören, da das Verständnis eines terminologischen Konzepts (insbesondere im Rahmen von terminologischen Taxonomien) eine Kenntnis auch des terminologischen Umfelds erfordert.

- (iii) Die Definitionsgebundenheit von Termini bedingt terminologische Diversität. Diese besteht i.a.R. nicht einfach zwischen einzelnen Termini, sondern zwischen ganzen terminologischen Systemen.

Die Freiheit zur Bedeutungsfestlegung, von der jeder Wissenschaftler durch die Ausführung definatorischer Gebrauchsfixierungshandlungen i.S.v. Wiegand (1996: 74) in seinen Beiträgen zum wissenschaftlichen Diskurs Gebrauch machen kann, impliziert stets auch die Freiheit aller anderen Akteure der Domäne, Gleiches zu tun. Sowohl die Prägung neuer als auch die form- oder begriffsseitige



Umprägung bereits eingeführter Termini ist dabei gebunden an sprachreflexives bzw. metasprachliches Handeln (vgl. Wiegand 1996; Budin 1996: 455f.). Im Unterschied zur Alltagssprache vollziehen sich die damit einhergehenden Veränderungen der in einer Fachdomäne verwendeten Lexik „nicht (sensu Keller 1995) von unsichtbarer Hand“ (Wiegand 1996: 95), sondern „aufgrund der gestalterischen Eingriffe sichtbarer Wissenschaftlerhände“.

Entsprechend sind Fachtermini im Kontext ihrer Verwendung (zumindest dann, wenn ihnen die explizite Formulierung einer eigenen Definition oder der Verweis auf die Definition eines anderen Urhebers vorausgegangen ist und der vom Autor gewählte Äußerungsplan von einem Rezipienten als Leseweg nachvollzogen wird) in aller Regel eindeutig. Mit Blick auf den fachlichen Diskurs der Domäne als Ganzer sind Termini aber oftmals ambig, da dieselben terminologischen Ausdrücke in unterschiedlichen Textäußerungen unterschiedlicher Autoren u.U. gemäß unterschiedlicher Definitionen verwendet werden. Terminologische Diversität tritt dabei in aller Regel nicht zufällig auf, sondern ist bewusst initiiert und Ausdruck eines Ringens um Konzepte und/oder um die Aktualisierung (Erweiterung, Eingrenzung) etablierter Konzepte in Abhängigkeit vom Erkenntnisfortschritt in Bezug auf den damit bezeichneten wissenschaftlichen Gegenstand:

Fachlicher Meinungsstreit findet wesentlich auf der Ebene der begrifflichen Klärung, Festlegung oder auch Uminterpretation von Fachwörtern statt. Auf diese Weise entsteht Polysemie, d. h., formal gleiche Fachwörter weisen verschiedene Bedeutungsvarianten auf. (Fraas 1997: 432)

Von der Formseite her betrachtet zeigt sich terminologische Diversität als Polysemie. Aus Sicht des einzelnen Wissenschaftsautors, der qua Definition (und häufig in bewusster Abgrenzung zu anderen Autoren) einen Terminus in seinem Sinne festlegt, handelt es sich hingegen um einen Akt der (Neu-)Konzeption eines wissenschaftlichen Gegenstandes. Der ggf. formgleiche Terminus eines anderen Autors wird dadurch nicht polysem; vielmehr wird dem bei einem anderen Autor mit dem betreffenden Ausdruck verbundenen Konzept (vorsätzlich) ein konkurrierendes Konzept gegenübergestellt, verbunden mit dem Anspruch, dass dieses dem damit zu beschreibenden Ausschnitt aus dem Objektbereich angemessener sei.

Dass ein Nebeneinander von verschiedenen Terminologisierungsvorschlägen für denselben Objektbereich wissenschaftlichen Interesses letztlich Wissenschaft wiederum voranbringen kann, ist den Experten der Domäne (= denjenigen, die Termini prägen und auf Basis ihrer Kenntnis des Objektbereichs und der auf ihn bezogenen Konzeptualisierungs- und Terminologisierungsvorschläge dazu in der Lage sind, kompetent Terminologiediskussionen zu führen) i.a.R. durchaus bewusst, wie das folgende Beispiel aus einer metalexikographischen Arbeit Herbert Ernst Wiegands beweist:

**Beispiel 3:** Textausschnitt aus Wiegand (2002: 413):

Es gibt derzeit keinen einheitlichen metalexikographischen Gebrauch von *Makrostruktur* und damit auch keinen einheitlichen Begriff von Makrostruktur innerhalb der Wörterbuchforschung. Nach der modernen Wissenschaftstheorie ist es keineswegs notwendig, daß ein wissenschaftliches Forschungsfeld nur über einen Begriff von einem bestimmten wissenschaftlichen Gegenstand verfügt. Ein Theorienpluralismus kann erkenntnisfördernd sein; er ist es vor allem dann, wenn die gegebenenfalls konkurrierenden theoretischen Begriffe als klar definiert gelten können [...].

Die Etablierung konkurrierender Terminologiesysteme ist damit aus Sicht der Experten nicht ein Ringen um die besten sprachlichen Etikettierungen von Konzepten, sondern in erster Linie ein Ringen um die beste (= plausibelste, erkenntnisförderlichste) konzeptuelle Gliederung eines Objektbereichs. Die jeweils unterschiedlichen Systeme werden von unterschiedlichen Gruppen innerhalb der Fachleute der betreffenden Domäne vertreten und propagiert. Mit Möhn (1997a: 177) können Fachsprachen als „Gruppensprachen von Experten“ und „intersubjektive Repräsentanten einer spezialisierten Wirklichkeitstopographie“ beschrieben werden, die eine ihrer wesentlichen Funktionen darin haben, eine „sprachlich manifestierte systematische Segmentierung und Differenzierung der Objektwelt“ (ebd.: 179) zu leisten und dabei – als Gruppensprache – „die einschlägigen Wirklichkeitsausschnitte (Gruppenkonzepte) kognitiv zu erfassen und auszudrücken“ (ebd.: 169) und für das Sprachhandeln der Mitglieder der Gruppe sprachlich verfügbar zu machen. Als Gruppensprache können Fachsprachen hierbei nicht lediglich auf der Ebene einzelner, z.B. institutionell repräsentierter Domänen (z.B. ‚Fachsprache der Linguistik‘, ‚Fachsprache der Physik‘) und ihrer Teildisziplinen, sondern unterhalb dieser Gliederungen auch mit Blick auf einzelne wissenschaftliche Paradigmen und Schulen beschrieben werden.

Für Novizen hingegen sind Termini zunächst Vokabeln, die sie von der Formseite her kennenlernen und die sie auf Konzept der Fachdomäne hinführen, wobei ihnen die konzeptuelle Gliederung des fachlichen Objektbereichs typischerweise erst in Ausschnitten bekannt ist. Der Objektbereich wird somit erschlossen über noch weitgehend unverbundene Ausschnitte aus dem terminologischen System der Domäne, die überdies nur einen von ggf. mehreren konkurrierenden Systemisierungsvorschlägen zum betreffenden Objektbereichsausschnitt darstellen können. Die Theorie- und Zweckabhängigkeit von Terminologie wird dabei oft (noch) nicht in vollem Umfang gewusst oder aber – da Terminologiereflexion zunächst eine solide Kenntnis des Objektbereichs und des zugehörigen fachlichen Diskurses voraussetzt – zunächst ausgeblendet, da ohne irgendeine grundlegende begriffliche Strukturierung der Objektbereich nicht durchdrungen und kognitiv in Relation zum eigenen Vorwissen gesetzt und an dieses angebunden werden kann. Entsprechend sind Novizen in aller Regel eher verunsichert, wenn sie bei ihrer primär auf die Gewinnung von Sachwissen orientierten Einarbeitung in eine Fachdomäne direkt mit terminologischer Pluralität und konkurrierenden Konzeptualisierungsvorschlägen konfrontiert werden. Der Erkenntnisgewinn durch konkurrierende „Weltentwürfe“ zu einzelnen Ausschnitten aus dem Objektbereich erschließt sich ihnen im Gegensatz zum Experten gerade *nicht* oder zumindest nicht ohne sorgfältige didaktische Aufbereitung. Folglich werden Abgrenzungsprobleme zwischen den Kategorien eines Systems oder alternative (ggf. feinere) Grenzziehungen zwischen Kategorien in einem alternativen System zum selben Objektbereich als Hürde beim Zugang zum Objektbereich empfunden, wie Storrer (2007) in Bezug auf die Vermittlung alternativer Wortarteneinteilungen in der Praxis der Lehraus- und -weiterbildung berichtet:

Gerade Studienanfänger erwarten häufig, dass „die Sprachwissenschaft“ (oder „der Duden“) eine feste und „richtige“ Menge an Wortarten vorgibt und festlegt, woran diese zu erkennen sind. Abgrenzungsprobleme und Grenzgänger werden dann als Mängel des betreffenden Wortartensystems bewertet. Die Tatsache, dass verschiedene Grammatiken des Deutschen sich in Zahl und Art der Wortarten unterscheiden, wird als verwirrend empfunden. Nach meiner

Erfahrung in der universitären Lehrerausbildung besteht im günstigen Fall der Wunsch, man möge als „Spezialistin“ festlegen, welches nun das „richtige“ Wortartensystem ist, und dieses erklären. Im ungünstigen Fall senkt die Frustration über die vermeintliche „Uneinigkeit“ der Grammatiker die Bereitschaft, sich überhaupt weiter mit dem Gebiet zu befassen (nach dem Motto: Wenn die Grammatiken sich nicht einigen können, welche und wie viele Wortarten es gibt, brauche ich das auch nicht zu wissen). (Ebd.: 906)

Um die Differenz zwischen konkurrierenden begrifflichen Systematisierungen zu einem Objektbereich als Gewinn, sprich: ihre Vor- und Nachteile als Anregung für eine genaue Prüfung der Systemisierungskriterien und damit für eine weitere Differenzierung des Blicks auf den systematisierten Gegenstand zu begreifen, muss mindestens eine Systematisierung zum Gegenstand bereits weitgehend vollständig bekannt und verstanden sein und muss ein grundlegendes Bewusstsein für Funktion und Methoden des terminologischen Kategorisierens von Weltausschnitten vorhanden sein. Semi-Experten, die i.d.R. bereits Experten in einer anderen Fachdomäne sind und somit ein wissenschaftstheoretisches Grundverständnis bereits mitbringen, fehlt es an ersterem; Novizen einer Fachdomäne (sofern sie nicht bereits in einer anderen Domäne zumindest Semi-Experte sind) fehlen in der Regel beide Voraussetzungen.

Neben der konkurrierenden terminologisch-konzeptuellen Gliederung von Objektbereichen sind sprachkontaktbedingte Benennungsalternativen eine weitere Quelle für terminologische Diversität: Insbesondere in Domänen, in denen Termini für grundlegende Konzepte zunächst im angloamerikanischen Diskurs geprägt wurden, findet sich in der deutschsprachigen Fachliteratur nicht selten ein Nebeneinander von entlehnten, ursprünglich englischen Termini und Lehnübersetzungen (vgl. Fraas 1997: 431). So finden sich beispielsweise in der Domäne „Hypermediaforschung und Texttechnologie“ unterschiedliche Vorschläge zur terminologischen Gliederung des Objektbereichs ‚Hyperlink‘. Während einige Autoren ihre Terminologie auf das Lehnwort *Link* aufbauen und die von ihnen unterschiedenen untergeordneten Konzepte mit analog gebildeten Bezeichnungen wie *1:1-Link*, *bidirektionaler Link* etc. versehen, arbeiten andere Autoren mit Lehnübersetzungen, z.B. Kuhlen (1991) mit dem Ausdruck *Verknüpfung* (und entsprechend *1:1-Verknüpfung*, *bidirektionale Verknüpfung* etc.) oder Tochtermann (1995) mit dem Ausdruck *Verweis* (und analog *1:1-Verweis*, *bidirektionaler Verweis* etc.) (vgl. Storrer 2010: 212). Selbst wenn einige dieser sprachkontaktbedingt verschieden benannten Termini eine große konzeptuelle Übereinstimmung aufweisen, können sie nicht einfach als Synonyme behandelt werden. Denn für sie ist charakteristisch, dass sie – da sie unterschiedlichen Terminologiesystemen und damit (autor-/schulenabhängig) unterschiedlichen Weisen des Redens über Hyperlinks zugehören – in beliebigen Texten der Fachdomäne gerade nicht füreinander ausgetauscht werden können.

Anforderung an die Modellierung: Eine Modellierung von Fachterminologie, die anstrebt, der Sprachwirklichkeit im Wissenschaftsdiskurs einer Domäne Rechnung zu tragen, muss in der Lage sein, sprachlich und konzeptuell unterschiedlich strukturierte Terminologiesysteme zu erfassen, als unterschiedlich zu beschreiben und konzeptuelle Bezüge zwischen ihnen aufzuzeigen.

Da Bezüge zwischen terminologischen Einheiten aus unterschiedlichen Terminologiesystemen in der Fachliteratur nicht selten explizit gemacht werden (häufig im

Umfeld von Terminologiediskussionen, die in die Einführung eigener neuer oder in die erneute Begründung bestimmter bereits etablierter Termini münden), in der Regel aber nie als 1:1-Zuordnungen dargestellt werden (da eine 1:1-Abbildbarkeit eines terminologischen Systems auf ein anderes den Sinn des jüngeren der beiden in Frage stellte), sind Zuordnungen zwischen Termini aus unterschiedlichen Systemen problematisch und schwierig begründbar. Im Rahmen von Überlegungen zum Vergleich konkurrierender Kategorieninventare zur Unterscheidung von Valenzpartnereinheiten hat Storrer (1992: 141) auf das wissenschaftstheoretische Problem der Inkommensurabilität von Konzepten hingewiesen, die unterschiedlichen theoretischen Entwürfen zum selben Objektbereich angehören; diese seien lediglich als kategorienähnlich aufzufassen, nicht aber als Entsprechungsbeziehungen zu beschreiben:

Die Zuordnung [...] wird dadurch erschwert, daß die [...] Kategorieninventare sich sowohl in der „Natur“ als auch in der Anzahl der berücksichtigten Kategorien unterscheiden. Von einem wissenschaftstheoretischen Standpunkt aus könnte sogar dafür argumentiert werden, daß die verschiedenen syntaktischen Kategorien inkommensurabel sind, weil sie zu verschiedenen Paradigmen wissenschaftlicher Beschreibung syntaktischer Strukturen gehören. Dem Problem der Vergleichbarkeit verschiedener Syntaxtheorien soll zumindest terminologisch Rechnung getragen werden, indem nicht von **Kategorienentsprechungs-** sondern lediglich von **Kategorienähnlichkeitsbeziehungen** gesprochen wird. Die Zuordnung sind also als Ähnlichkeitsbeziehungen zu verstehen, die nur für den vorliegenden Zweck aufgestellt worden sind; aus wissenschaftstheoretischer Sicht müßte eine derartige Zuordnung wesentlich ausführlicher problematisiert und begründet werden.

Im TermNet-Ansatz, der im folgenden Abschnitt vorgestellt wird, spielt die Unterscheidung zwischen **Kategorienentsprechung** und **Kategorienähnlichkeit** eine zentrale Rolle. **Kategorienähnlichkeit** wird entsprechend zu Storrer (1992) als ein zweckmäßiges Konstrukt aufgefasst, das es erlaubt, Termini übergreifend zu den terminologischen Systemen, in die sie eingebunden sind und in deren Rahmen sie ihre spezifische Funktion haben, lose darüber zueinander in Beziehung zu setzen, dass sie wissenschaftliche Gegenstände mit ähnlichen Eigenschaften konstituieren bzw. einen ähnlichen Ausschnitt aus dem wissenschaftlichen Objektbereich beschreiben (dies aber im Detail auf unterschiedliche Weise, z.B. unter unterschiedlichem Aspekt oder in partiell unterschiedlichem Zuschnitt, leisten).

### 3. Der *TermNet*-Ansatz

#### 3.1 Projektkontext und Anwendungsszenario

In den beiden Projektphasen des HyTex-Projekts wurden unter dem Namen *TermNet* zwei verschiedene Versionen eines terminologischen Wortnetzes zur Fachdomäne „Hypermediaforschung und Texttechnologie“ entwickelt. Die Modellierung der beiden Versionen unterscheidet sich geringfügig, folgt aber derselben Grundidee, die in diesem Abschnitt in ihren wesentlichen Zügen vorgestellt werden soll.

Die in Projektphase I (2001–2005) aufgebaute Ressource umfasst 226 Termini und wurde komplett im ISO-Standard XML Topic Maps (XTM; vgl. Pepper & Moore 2001) repräsentiert. Die Ressource aus Projektphase II (2005–2009) umfasst 423 Termini und wurde in verschiedenen Versionen der Ontologiebeschreibungssprache OWL (Web Ontology Language; Dean & Schreiber 2004) abgebildet und mit einer OWL-Repräsentation eines Ausschnitts des allgemeinsprachlichen deutschen Wortnetzes *GermaNet* verknüpft (*GermaTermNet*; Kunze et al. 2007; Lungen & Storrer 2007; Lungen et al. 2008). Die Repräsentation der Modellierung in XTM bzw. OWL sowie die dabei getroffenen Entscheidungen sind nicht Gegenstand der nachfolgenden Darstellung. Die XTM-Repräsentation ist im Detail beschrieben in Lenz, Beißwenger & Storrer (2002) und Lenz, Birkenhake & Maas (2004), zur Modellierung in OWL vgl. Selzam (2009) und Lungen et al. (i. Dr.). Die verschiedenen OWL-Repräsentationen von *TermNet-II* sowie der integrierten *GermaTermNet*-Ressource können online unter <http://www.wordnets-in-owl.de> heruntergeladen werden.

Im Fokus der folgenden Skizze stehen die grundlegenden Modellierungsentscheidungen und -konstrukte in *TermNet*. Im Detail kann die Modellierung für *TermNet-I* in Beißwenger, Storrer & Runte (2004) und Runte (2004) sowie für *TermNet-II* anhand von Beißwenger (2008) nachvollzogen werden.

Die für Einträge in der *TermNet*-Ressource in Frage kommenden Fachtermini wurden korpusgestützt ermittelt: Aufgenommen wurden zunächst solche Termini, für die bei einer intellektuellen Sichtung des zugrunde gelegten Fachtextkorpus das Vorhandensein vertexteter Definitionen festgestellt werden konnte. Für die Erweiterung der *TermNet*-Ressource in Projektphase II wurde der Bestand an terminologischen Einträgen für einzelne Terminologiebereiche aus Konsistenzgründen um Einträge zu weiteren Termini ergänzt, die im zugrunde liegenden Korpus zwar nicht mit Definitionen belegt waren, die nach Einschätzung zweier Expertinnen aber weitere relevante Konzepte der Domäne repräsentierten.

Das zugrunde gelegte Fachtextkorpus ist textsortenheterogen und umfasst neben wissenschaftlichen Fachartikeln auch Dokumenttypen mit normativem („Spezifikation“) und didaktischem Anspruch („Tutorial“, „Einführung“, „Überblicksdarstellung“), des weiteren diskursiv geprägte Textsorten wie FAQs, Mailinglisten- oder Foren-Postings. Eine Dokumentation zu Inhalt und Zusammensetzung des Korpus findet sich in Beißwenger & Wellinghoff (2003/2006). Ein ausgezeichnete Ausschnitt aus dem Gesamtkorpus – im Projektkontext als *Kernkorpus* bezeichnet – wurde auf mehreren Annotationsschichten linguistisch annotiert (Storrer 2010). Für die Vernetzung der *TermNet*-Ressource mit dem Fachtextkorpus war hierbei insbesondere die Annotationsschicht „Definitionen und Termverwendungsinstanzen“ relevant, auf welcher (ma-

nuell) alle in den Texten enthaltenen definitorischen Textsegmente sowie (semi-automatisch) sämtliche Verwendungsinstanzen der betreffenden Termini ausgezeichnet wurden (vgl. Beißwenger 2004, Lenz, Beißwenger & Wellinghoff 2004/2006).

### 3.2 Modellierung des terminologischen Wortnetzes

TermNet ist ein Wortnetz im Stile des Princeton WordNet (Fellbaum 1998). Beim Aufbau der Ressource ging es nicht darum (wie etwa im brasilianischen TermiNet-Projekt, vgl. Di Felippo 2010), fachsprachlichen Wortschatz mit dem WordNet-Modell zu erfassen oder (wie z.B. in Magnini & Speranza 2002 und Kunze et al. 2007) fachsprachliche Ressourcen mit gemeinsprachlichen Wortnetzen zu verknüpfen. Statt dessen war es das Ziel, eine Modellierung zu finden, die auf die WordNet-Modellierung aufbaut und diese in einer Weise erweitert, dass die Besonderheiten fachsprachlicher Wortschätze und insbesondere terminologische Diversität damit erfasst und abgebildet werden können.

Die beiden grundlegende Modellierungseinheiten in Wortnetzen heißen *words* und *synsets*: *words* repräsentieren Wortbedeutungen bzw. Lesarten, *Synsets* (= „sets of synonyms“) repräsentieren Mengen synonyme Wortbedeutungen und damit – indirekt – Konzepte. Die Beziehung zwischen einer Wortbedeutung und dem Synset, dem sie zugeordnet ist, wird als *Lexikalisierungsbeziehung* aufgefasst; zwei Wortbedeutungen, die demselben Synset angehören, stehen zueinander in einer *Synonymiebeziehung*. Für die Zusammenfassung von Wortbedeutungen zu Synsets wird dabei ein weit gefasstes Synonymieverständnis zugrunde gelegt („Austauschbarkeit nicht in allen, sondern in einigen Kontexten“, vgl. Miller 1998: 23f.). Die Beziehungen zwischen Wortbedeutungen werden durch eine Reihe von lexikalischen Relationen, die Beziehungen zwischen Synsets durch eine Reihe konzeptueller Relationen beschrieben (vgl. z.B. Kunze 2001).

Die beiden grundlegenden Einheiten von Wortnetzen – Wortbedeutungen (Lesarten) und Synsets – haben in TermNet ihre Entsprechung in den Einheiten *Terminus* und *Termset*. *Termini* sind Einheiten, denen im Rahmen eines bestimmten terminologischen Systems eine (qua Definition) genau fixierte Bedeutung zukommt und in denen sich ein terminologischer Ausdruck exklusiv mit einem bestimmten Konzept verbindet. Auf der Ebene der Zuordnung von *Termini* (aus unterschiedlichen Systemen) zu *Termsets* adaptiert der TermNet-Ansatz das Konzept der *Synsets*, reinterpretiert diese aber als Modellierungskonstrukte zur Darstellung von *Kategorienähnlichkeit* (zwischen terminologischen Konzepten aus unterschiedlichen Systemen) anstatt zur Darstellung von *Synonymie* (zwischen Wortbedeutungen). Die *Lexikalisierungsbeziehung*, die im WordNet-Ansatz zwischen Wortbedeutungen und *Synsets* besteht, besteht auf der Ebene einzelner terminologischer Systeme zwischen einem terminologischen Ausdruck und dem ihm definitorisch zugeordneten terminologischen Konzept. Die Beziehung zwischen *Termini* und *Termsets* wird als *Konzeptualisierungsbeziehung* aufgefasst. *Termini* (aus unterschiedlichen Systemen), die demselben *Termset* zugeordnet sind, stehen (implizit) in einer Beziehung der *Kategorienähnlichkeit*. Hierarchische Beziehungen zwischen *Termini* innerhalb desselben terminologischen Systems und Beziehungen zwischen *Termsets* werden aufgrund ihres unterschiedlichen Status anhand unterschiedlicher Relationen beschrieben.

Dem TermNet-Modell liegen u.a. die folgenden Überlegungen und Entscheidungen zugrunde:

- (i) Innerhalb eines terminologischen Systems entspricht ein Terminus i.a.R. genau einem Konzept. Wir verzichten daher auf die Unterscheidung zwischen Termini und Synsets und fassen Termini als Einheiten auf, in denen eine (fach-)lexikalische Einheit genau ein Konzept lexikalisiert und ein- und dasselbe Konzept nur durch einen terminologischen Ausdruck lexikalisiert ist. Fälle, in denen synonyme Termini innerhalb eines Systems auftreten, betrachten wir als wortbildungsmorphologisch erklärbar. Fälle dieser Art (z.B. das Vorhandensein der synonymen Termini ‚Dokumenttyp-Definition‘ und ‚DTD‘ oder ‚Hypertextmodul‘ und ‚Modul‘) behandeln wir als zwei Termini, weisen sie aber durch Etablierung einer entsprechenden Relation (‚ist Akronym zu‘ mit der Konverse ‚ist Vollform zu‘ oder ‚ist Abkürzung zu‘ mit der Konverse ‚ist Expansionsform zu‘) als terminologische Dubletten aus.<sup>1</sup>
- (ii) Die Modellierungseinheit ‚Termset‘ ist ein zweckmäßiges Konstrukt zur Beschreibung von Bezügen zwischen einzelnen terminologischen Systemen. Die Einführung der Einheit ‚Termset‘ ist auf den Zweck der Modellierung bezogen (= eine Ressource aufzubauen, anhand derer sich Semi-Experten und Novizen Orientierung über Ähnlichkeiten zwischen Termini aus unterschiedlichen terminologischen Systemen verschaffen können). Sie beansprucht keinen ontologischen Status, sondern repräsentiert einen Teil des Expertenwissens zur Domäne – nämlich denjenigen, der es Experten erlaubt, alternative terminologische Konzeptualisierungssysteme zu denselben Ausschnitten aus dem fachlichen Objektbereich zueinander in Beziehung zu setzen und zu vergleichen. Die Zuordnung von Termini aus unterschiedlichen Systemen zu demselben Termset im TermNet-Modell stellt notwendigerweise eine Abstraktion von den Feinheiten der einzelnen Terminologiesysteme dar. Im Detail können die Unterschiede zwischen konkurrierenden terminologischen Systemen am besten durch ein Studium der Fachtexte nachvollzogen werden. Die TermNet-Ressource beansprucht nicht, die Textlektüre zu ersetzen. Statt dessen zielt die Einführung des Termset-Konstrukts darauf, Semi-Experten und Novizen einen „grobkörnigeren“ Zugang zum Objektbereich der Domäne zu eröffnen, der es erlaubt, zunächst die Gemeinsamkeiten zwischen alternativen terminologischen Systemen zu erschlie-

---

<sup>1</sup> In der Modellierung von TermNet-I haben wir (in enger Anlehnung an die WordNet-Modellierung) Termini ähnlicher Bedeutung aus unterschiedlichen terminologischen Systemen noch als Lexikalisierungen desselben Termsets aufgefasst und das Konzept der ‚Termsets‘ analog zum Konzept der ‚Synsets‘ gefasst. Entsprechend stellte sich die Notwendigkeit zur Beschreibung terminologischer Synonymie weit- aus häufiger (z.B. waren dann auch sprachkontaktbedingte Varianten – etwa das Vorhandensein von Entlehnungen und Lehnübersetzungen zu einem weitgehend identischen terminologischen Konzept – in der Beschreibung zu berücksichtigen, obwohl diese innerhalb desselben terminologischen Systems i.d.R. nicht auftreten) (vgl. Beißwenger, Runte & Storrer 2004; Beißwenger 2008). In TermNet-II haben wir Beziehungen innerhalb desselben terminologischen Systems und Bezüge zwischen Einheiten aus unterschiedlichen terminologischen Systemen anhand verschiedener Arten von Relationen beschrieben und Termsets – wie dargelegt – ausschließlich als Konstrukte zur Beschreibung von Kategorienähnlichkeit zwischen terminologischen Einheiten aus unterschiedlichen Systemen genutzt. Dies erlaubte es zum einen, unterschiedliche Terminologiesysteme in der Modellierung konsequenter voneinander zu unterscheiden und reduzierte zum anderen die bei der Modellierung zu behandelnden Fälle von Synonymie innerhalb ein- und desselben terminologischen Systems auf wortbildungsmorphologische Varianten.

ßen. Durch die Modellierung verschiedener terminologischer Systeme *und* deren Vernetzung (*cum grano salis*) anhand von Termsets erachten wie es gerade als eine Stärke des Ansatzes, unterschiedlich granulare Zugänge zur terminologischen Gliederung des Objektbereichs einer Domäne zu eröffnen: Der Semi-Experte/Novize kann sich zunächst termset-geleitet einen groben Überblick über die Gegenstände der Domäne verschaffen, um sich anschließend – vermittelt über die Termset–Termini-Beziehungen – im Detail über einzelne terminologische Systeme und deren Zuschnitt zu informieren. Er kann aber auch – von der Lektüre eines bestimmten Textes herkommend – sich zunächst das terminologische System des aktuell gelesenen Autors erschließen, bevor er sich dann in einem Abstraktionsschritt – vermittelt über die Termini–Termset-Beziehungen – über dessen Bezüge zu alternativen terminologischen Systemen orientiert.

- (iii) Die hierarchischen Beziehungen zwischen Termini desselben terminologischen Systems und zwischen Termsets haben einen unterschiedlichen Status und werden anhand unterschiedlicher Relationstypen beschrieben. Wir gehen davon aus, dass ein terminologisches System immer eine bestimmte Sichtweise (Ideologie) in Bezug auf den zugeordneten Objektbereich repräsentiert und dass Beziehungen zwischen Termini desselben terminologischen Systems stets in irgendeiner Weise in Textäußerungen der Fachdomäne dokumentiert sind (z.B. dadurch, dass das übergeordnete terminologische Konzept als *genus proximum* im Definiens einer Definition zu einem untergeordneten terminologischen Konzept erwähnt ist).

Die Beschreibung von Termsets ist hingegen didaktisch motiviert; in der Praxis des Wissenschaftsdiskurses, so wie er sich in Fachtexten der Domäne zeigt, treten sie als wissenschaftliche Gegenstände nicht in Erscheinung (wissenschaftliche Gegenstände existieren vielmehr erst unter der Perspektive konkreter terminologischer Systeme in Form einzelner terminologischer Konzepte). Zwar werden in Fachtexten bisweilen konkurrierende terminologische Konzepte, die ähnliche Ausschnitte aus dem Objektbereich in je unterschiedlicher Weise als wissenschaftlichen Gegenstände konstituieren, miteinander verglichen – die Termsets selbst (die *cum grano salis* Ausschnitte aus dem Objektbereich repräsentieren) sind jedoch noch kein wissenschaftlicher Gegenstand. Entsprechend werden die Beziehungen zwischen ihnen erst bei der Modellierung hergestellt (auf Basis der Kompetenz von an der Modellierung beteiligten Experten der Domäne, Kategorienähnlichkeiten zwischen alternativen terminologischen Konzepten zu erkennen).

Innerhalb einzelner Terminologiesysteme sind Beziehungen der Über-/Unterordnung zwischen terminologischen Konzepten i.d.R. als Klasseninklusion konzipiert. Entsprechend werden sie im TermNet-Ansatz mit den Thesaurus-Relationen ‚Oberbegriff‘/‚Unterbegriff‘ dargestellt, die sich als Relationen zwischen Extensionen auffassen lassen, bei denen die übergeordnete Klasse die untergeordnete vollständig inkludiert. Die Beziehungen zwischen Termsets werden demgegenüber mit den aus dem WordNet-Ansatz bekannten Relationen ‚Hyponymie‘/‚Hyperonymie‘ beschrieben. Diese lassen sich – z.B. bei der formalen Repräsentation des Terminologiemodells in einer Ontologie – im Bedarfsfall als Klasseninklusion reinterpretieren, sind aber prinzipiell als „weiche“ Relationen zwischen Konzepten konzipiert und daher besser geeignet, Beziehungen zwischen



Mengen von kategorienähnlichen Konzepten zu beschreiben. Daneben werden auf der Ebene der Termsets Meronymie- und Holonymie-Beziehungen dargestellt.

### 3.3 Zwei Anwendungsbeispiele

Im Rahmen des HyTex-Projekts wurden in einem Demonstrations-Prototypen die im Projekt entwickelten Strategien für die Hypertextualisierung nach Wissensvoraussetzungen testweise umgesetzt. Der Prototyp, der von der Projektwebsite (<http://www.hytex.info>) heruntergeladen werden kann, simuliert eine Hypertext-Leseumgebung für ausgewählte Fachtexte der Domäne „Hypermediaforschung und Texttechnologie“. Das terminologische Wortnetz TermNet (in der Version aus Projektphase I) ist in die Leseumgebung in Form einer Glossars integriert, das kategorienähnliche Termini zu Termsets gruppiert und mit anderen Termsets in Beziehung setzt. Abb. 3 zeigt die Termset-Darstellung, die beim Nachschlagen des Terminus ‚Knoten‘ angezeigt wird; ein Nachschlagen der demselben Termset zugeordneten Termini ‚Hypertextmodul‘, ‚informationelle Einheit‘, ‚Modul‘ etc. führt – unter entsprechend angepasster Überschrift – zu einer identischen Darstellung. Dem Eintrag nebengeordnet sind Hyperlinks auf Definitionen des Terminus in den in der Leseumgebung zur Verfügung stehenden, hypertextuell präsentierten Fachtexten (Nr. 1 in der Abbildung). Die Definitionen können aus dem Glossar heraus aufgerufen werden (2), wahlweise können sie aber auch im Kontext der Original-Textstelle eingesehen werden (3). Das Glossar führt auf diese Weise über die für einen Terminus in den Texten belegten Definitionen in die Texte hinein. Umgekehrt können bei der Textrezeption über Hyperlinks, die Verwendungsinstanzen von Termini im Text mit den zugehörigen Glossareinträgen vernetzen, die jeweils zugehörigen Glossareinträge aufgerufen und eingesehen werden (4). Das Glossar kann somit sowohl zur Erarbeitung von Wissen über die Strukturiertheit der in der Domäne verwendeten Terminologie (lernbezogen) als auch zur Unterstützung bei der Textrezeption (textbezogen) verwendet werden.

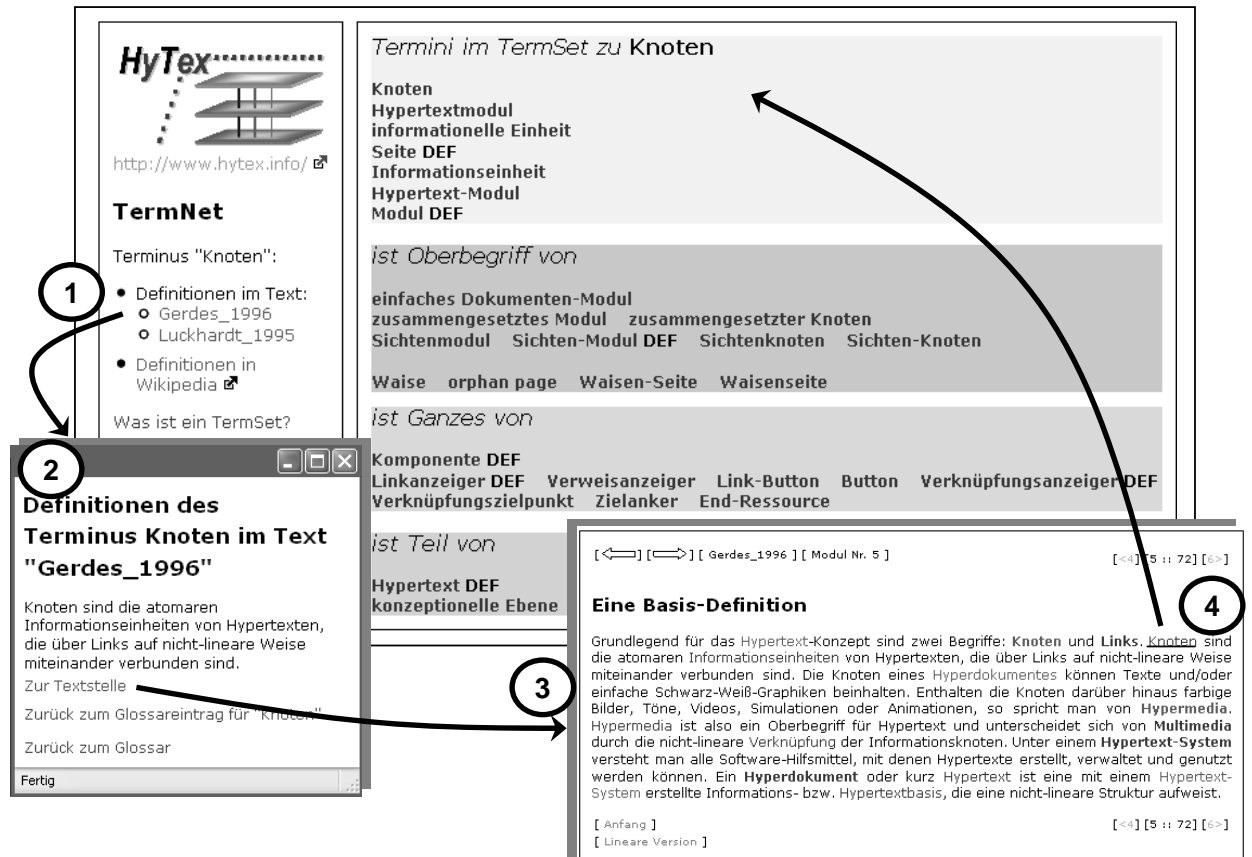


Abb. 3: Darstellung des Termsets ‚Knoten‘ in der Glossaranwendung des HyTex-Prototypen.

Die im HyTex-TermNet umgesetzte Idee, kategorienähnliche terminologische Konzepte aus unterschiedlichen terminologischen Systemen über ein Termset-Konstrukt zueinander in Beziehung zu setzen und auf diese Weise Bezüge zwischen terminologischen Gliederungsentwürfen zum selben Objektbereich zu verdeutlichen, wurde auch in der Komponente „Grammatische Ontologie“ des hypertextuell organisierten grammatischen Informationssystems Grammis (<http://hypermedia.ids-mannheim.de/grammis/>) am Institut für deutsche Sprache (IDS) Mannheim übernommen (vgl. Schneider 2007). In dieser Ontologie sind grammatische Termini zu Termsets gruppiert. Abb. 4 zeigt den Eintrag, der beim Nachschlagen nach einem der Termini ‚Pronomen‘, ‚Proterm‘, ‚Fürwort‘, ‚Proform‘, ‚Pronominalwort‘ oder ‚Stellvertreter‘ angezeigt wird. Die Termini ‚Pronomen‘ und ‚Proterm‘ werden dabei als von den Autoren der Ontologie favorisierte Termini gekennzeichnet. Dem Termset sind weitere Termsets über- und untergeordnet, die zu ihm in Hyperonymie-, Hyponymie- und Meronymiebeziehungen stehen. Angezeigt wird von diesen Termsets zunächst nur der favorisierte Terminus (z.B. ‚Personalpronomen‘), per Mausklick auf den jeweiligen Kasten wird das betreffende Termset ins Zentrum der Darstellung gerückt und werden sämtliche darin enthaltenen Termini angezeigt (z.B. ‚Personalpronomen‘, ‚persönliches Fürwort‘). Dem zentral dargestellten Termset ist jeweils ein Buchsymbol nebengeordnet, das zum Aufruf eines Eintrags in der Grammis-Komponente „Grammatische Fachbegriffe“ führt, der Detailinformationen zum terminologischen Konzept bietet. Von diesem Eintrag können über weitere Linkangebote direkt diejenigen Passagen aus der Grammis-Komponente „Systematische Grammatik“ aufgerufen werden, in denen das Konzept und der zugehörige Ausschnitt aus dem fachlichen Objektbereich ausführlich dargestellt sind.

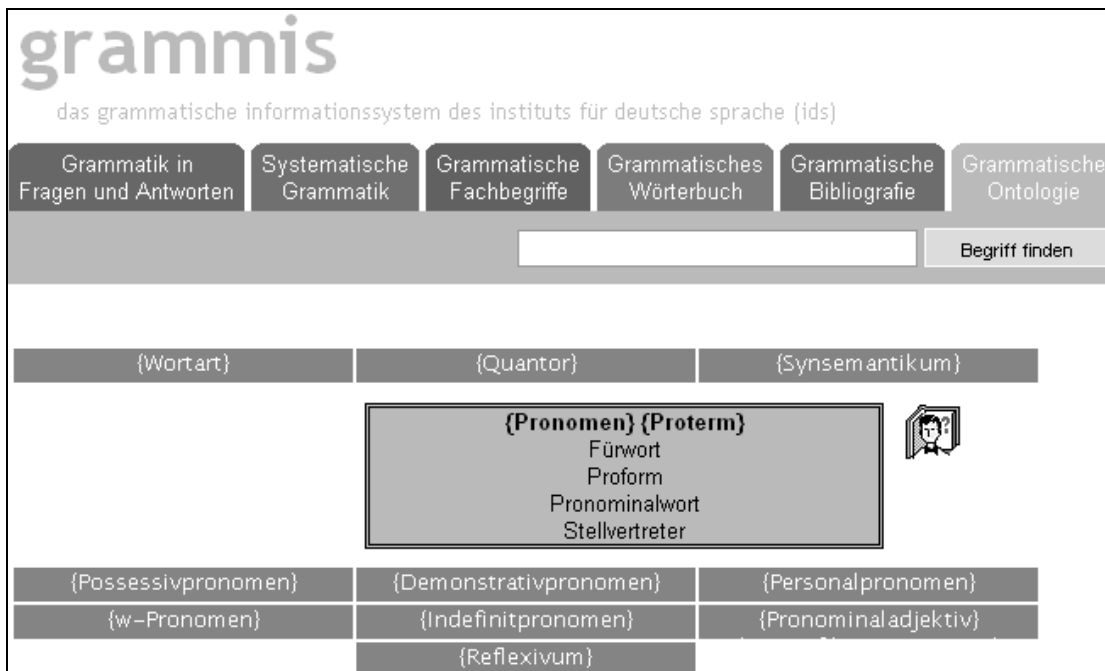


Abb. 4: Darstellung des Termsets ‚Pronomen/Proterm‘ in der „Grammatischen Ontologie“ in Grammis.

#### 4. Fazit

Der vorgestellte Ansatz setzt alle drei der in Abschnitt 2 formulierten Anforderungen um und erlaubt es, nicht nur ein terminologisches System als in sich konsistenten sprachlich-konzeptuellen Entwurf zu einem fachlichen Objektbereich darzustellen, sondern mehrere alternative Systeme in einem Modell abzubilden. Das Konstrukt des Termsets erlaubt es, Kategorienähnlichkeiten zwischen Einheiten aus unterschiedlichen Systemen aufzuzeigen und im Modell zu erfassen. Damit wird zwar auf der Ebene der Termsets von bestimmten (für ein genaues Verständnis einzelner terminologischer Systeme wichtigen) Feinheiten abstrahiert; andererseits erlaubt es die Rückbindung der terminologischen Einträge im Wortnetz an relevante Passagen in der zugrunde liegenden Textbasis, den konzeptuellen Zuschnitt der betreffenden Termini (als Einheiten im Rahmen einzelner Systeme) anhand einschlägiger Textäußerungen der Fachdomäne nachzuvollziehen und sich Detailwissen zu ihren systematischen Beziehungen zu anderen Systemeinheiten zu erarbeiten. Ressourcen wie das im HyTex-Projekt aufgebaute TermNet oder die „Grammatische Ontologie“ in Grammis ermöglichen somit die Erarbeitung terminologischen Wissens auf zwei Granularitätsstufen: einmal auf Ebene einzelner Systeme, und einmal anhand einer groben Strukturierung der in der Domäne verhandelten Gegenstände (die aber letztlich immer wieder auf einzelne terminologische Systeme zurückführt).

Durch die systematische Vernetzung des Terminologiemodells mit der zugrundeliegenden digitalen Textbasis ist das im Modell repräsentierte terminologische Wissen dabei stets rückgebunden an diejenigen wissenschaftlichen Textäußerungen, in denen es im fachlichen Diskurs der Domäne originär verankert ist. Ressourcen im Stile des TermNet oder der „Grammatischen Ontologie“ (die speziell für Semi-Experten und Novizen der jeweiligen Domänen konzipiert sind) ersetzen somit nicht die vertiefte Auseinanderset-

zungen mit Fachtexten, sondern führen die Nutzer der auf ihrer Basis bereitgestellten digitalen Nachschlagewerke letztlich immer wieder in die Fachtexte hinein. Bei der textbezogenen Nutzung ermöglichen sie die Erarbeitung von Wissen zum semantisch-konzeptuellen Umfeld eines bei der Textrezeption vorgefundenen Terminus und seine Anbindung an bereits gekannte Termini aus konkurrierenden Systemen. Bei der Nutzung zur Erarbeitung von Wissen über die in der Domäne relevanten fachlichen Gegenstände orientieren sie über existierende terminologische Konzepte zum Objektbereich (inklusive Alternativen) und ermöglichen einen konzeptgeleiteten Zugang zur einschlägigen Fachliteratur.

## 5. Literatur

- Beißwenger, Michael (2004): Annotation definitorischer Textsegmente und terminologiesensitives Linking. Interner Report der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“. <http://coli.lili.uni-bielefeld.de/Texttechnologie/Forschergruppe/publications/hytex-deflink.pdf> (10.8.2010).
- Beißwenger, Michael (2008): TermNet – ein terminologisches Wortnetz im Stile des Princeton WordNet. Interner Report der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“. <http://coli.lili.uni-bielefeld.de/Texttechnologie/Forschergruppe/publications/hytex-termnet2008.pdf> (10.8.2010).
- Beißwenger, Michael, Angelika Storrer & Maren Runte (2004): Modellierung eines Terminologienetzes für das automatische Linking auf der Grundlage von WordNet. In: Kunze, Lemnitzer & Wagner (Hrsg.), 113-126.
- Beißwenger, Michael & Sandra Wellinghoff (2003/2006): Inhalt und Zusammensetzung des Fachtextkorpus. Interner Report der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“. <http://coli.lili.uni-bielefeld.de/Texttechnologie/Forschergruppe/publications/hytex-korpuszusammensetzung.pdf> (10.8.2010).
- Bergenholtz, Henning & Uwe Kaufmann (1997): Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. In: *Hermes* 18, 91-125.
- Budin, Gerhard (1996): Wissensorganisation und Terminologie. Die Komplexität und Dynamik wissenschaftlicher Informations- und Kommunikationsprozesse. Tübingen (Forum für Fachsprachen-Forschung 28).
- Dean, Mike & Guus Schreiber (Eds., 2004): OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004. WWW-Ressource: <http://www.w3.org/TR/owl-ref/>
- Di Felippo, Ariani (2010): The TerminiNet Project: an Overview. In: Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas. Los Angeles, 92-99. <http://www.aclweb.org/anthology/W/W10/W10-1613.pdf> (10.8.2010).
- Fellbaum, Christiane (Ed., 1998): WordNet. An Electronic Lexical Database. Cambridge/Mass. (Language, Speech, and Communication).
- Fluck, Hans-R. (1996): Fachsprachen. Einführung und Bibliographie. 5., überarb. u. erw. Aufl. Tübingen. Basel (UTB 483).
- Fraas, Claudia (1997): Lexikalisch-semantische Eigenschaften von Fachsprachen. In: HSK 14.1, 428-438.
- HSK 3.1 = Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. Hrsg. v. Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand, Ladislav Zgusta. 1. Teilbd. Berlin. New York 1989 (Handbücher zur Sprach- und Kommunikationswissenschaft 3.1).

- HSK 14.1 = Fachsprachen. Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research / Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. Hrsg. v. Lothar Hoffmann, Hartwig Kalverkämper, Herbert Ernst Wiegand, in Verbindung mit Christian Galinski und Werner Hüllen. 1. Teilbd. Berlin. New York 1997 (Handbücher zur Sprach- und Kommunikationswissenschaft 14.1).
- HSK 14.2 = Fachsprachen. Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research / Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. Hrsg. v. Lothar Hoffmann, Hartwig Kalverkämper, Herbert Ernst Wiegand, in Verbindung mit Christian Galinski und Werner Hüllen. 2. Teilbd. Berlin. New York 1998 (Handbücher zur Sprach- und Kommunikationswissenschaft 14.2).
- Kuhlen, Rainer (1991): Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank. Berlin etc.
- Kunze, Claudia (2001): Lexikalisch-semantische Wortnetze. In: Kai-Uwe Carstensen, Christian Ebert, Cornelia Endriss, Susanne Jekat, Ralf Klabunde & Hagen Langer (Hrsg.): *Computeringuistik und Sprachtechnologie. Eine Einführung*. Heidelberg. Berlin, 386-393.
- Kunze, Claudia & Andreas Wagner (2001): Anwendungsperspektiven des GermaNet, eines lexikalisch-semantischen Netzes für das Deutsche. In: Ingrid Lemberg, Bernhard Schröder & Angelika Storrer (Hrsg.): *Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*. Tübingen (Lexicographica. Series Maior 107), 229-246.
- Kunze, Claudia, Lothar Lemnitzer, Harald Lungen & Angelika Storrer (2007): Repräsentation und Verknüpfung allgemeinsprachlicher und terminologischer Wortnetze in OWL. In: *Zeitschrift für Sprachwissenschaft* 26, 267-290.
- Kunze, Claudia, Lothar Lemnitzer & Andreas Wagner (Hrsg., 2004): *Anwendungen des Deutschen Wortnetzes in Theorie und Praxis. Beiträge des GermaNet-Workshops Tübingen, Oktober 2003 (LDV-Forum 19 (1/2))*.
- Lenz, Eva Anna, Michael Beißwenger & Angelika Storrer (2002): Hypertextualisierung mit Topic Maps - ein Ansatz zur Unterstützung des Textverständnisses bei der selektiven Rezeption von Fachtexten. In: Robert Tolksdorf & Rainer Eckstein (Hrsg.): *XML Technologien für das Semantic Web (XSW 2002). Proceedings zum Workshop 24.-25. Juni 2002, Berlin*. Bonn: Gesellschaft für Informatik (Lecture Notes in Informatics), 151-159.
- Lenz, Eva Anna, Michael Beißwenger & Sandra Wellinghoff (2004/2006): Annotationsschicht: Definitionen und Termverwendungsinstanzen. Interner Report der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“. <http://coli.lili.uni-bielefeld.de/Texttechnologie/Forschergruppe/publications/hytext-annotation-termini-und-definitionen.pdf> (10.8.2010).
- Lenz, Eva Anna, Benjamin Birkenhake, Jan Frederik Maas (2004): Von der Erstellung bis zur Nutzung: Wortnetze als XML Topic Maps. In: Kunze, Lemnitzer & Wagner (Hrsg.), 127-136.
- Lungen, Harald, Claudia Kunze, Lothar Lemnitzer & Angelika Storrer (2008): Towards an Integrated OWL Model for Domain-Specific and General Language WordNets. In: *Proceedings of the 4th Global Word-Net Conference*. 22.-25.01.2008, Szeged, Ungarn. 281-296. [http://www.inf.u-szeged.hu/projectdirs/gwc2008/GWC2008\\_Proceedings\\_Final.pdf](http://www.inf.u-szeged.hu/projectdirs/gwc2008/GWC2008_Proceedings_Final.pdf) (10.8.2010).
- Lungen, Harald & Angelika Storrer (2007): Domain Ontologies and Wordnets in OWL: Modeling Options. In: Uwe Mönnich & Kai-Uwe Kühnberger (Eds.): *Foundations of Ontologies in Text Technology*. (LDV-Forum 22(2)), 1-17.

- Lüngen, Harald, Michael Beißwenger, Bianca Selzam & Angelika Storrer (im Druck): Modeling and Processing Wordnets in OWL. In: Alexander Mehler, Kai-Uwe Kühnberger, Hennig Lobin, Harald Lüngen, Angelika Storrer & Andreas Witt (Eds.): Modeling, Learning and Processing of Text Technological Data Structures. Dordrecht.
- Magnini, Bernardo & Manuela Speranza (2002): Merging Global and Specialized Linguistic Ontologies. In: Proceedings of Ontolex 2002. Las Palmas de Gran Canaria, 43–48.
- Miller, George A. (1998): Nouns in WordNet. In: Fellbaum (Ed.), 23-46.
- Möhn, Dieter (1997): Fachsprache als Gruppensprache. In: HSK 14.1, 150–157.
- Möhn, Dieter (1997a): Fachsprache und Gruppensprachen. In: HSK 14.1, 168–181.
- Pepper, Steve & Graham Moore (Eds., 2001): XML Topic Maps (XTM) 1.0. TopicMaps.Org Specification. WWW-Ressource: <http://www.topicmaps.org/xtm/1.0/> (10.8.2010).
- Roelcke, Thorsten (2005): Fachsprachen. 2., durchges. Aufl. Berlin (Grundlagen der Germanistik 27).
- Runte, Maren (2004): Modellierung des Terminologischen Wissensnetzes TermNet. Mit einem Anhang: Der WordNet-Ansatz – ein Überblick. Interner Report der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“. <http://coli.lili.uni-bielefeld.de/Texttechnologie/Forschergruppe/publications/hytex-termnet2004.pdf> (10.8.2010).
- Schneider, Roman (2007): A Database-driven Ontology for German Grammar. In: Georg Rehm, Andreas Witt & Lothar Lemnitzer (Hrsg.): Datenstrukturen für linguistische Ressourcen und ihre Anwendungen. Data Structures for Linguistic Resources and Applications. Proceedings of the Biennial GLDV Conference 2007. Tübingen, 305-314.
- Selzam, Bianca (2009): TermNet und GermaTermNet in OWL. Interner Report der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“. [http://www.hytex.tu-dortmund.de/pdf/Selzam\\_2009\\_TermNet-und-GermaTermNet-in-OWL.pdf](http://www.hytex.tu-dortmund.de/pdf/Selzam_2009_TermNet-und-GermaTermNet-in-OWL.pdf) (10.8.2010).
- Storrer, Angelika (1992): Verbvalenz. Theoretische und methodische Grundlagen ihrer Beschreibung in Grammatikographie und Lexikographie. Tübingen (RGL 126).
- Storrer, Angelika (2007): Grenzgänger: Problemfelder aus didaktischer Sicht. In: Ludger Hoffmann (Hrsg.): Handbuch der deutschen Wortarten. Berlin. New York, 905-924.
- Storrer, Angelika (2010): Mark-up Driven Strategies for Text-to-Hypertext Conversion. In: Dieter Metzger & Andreas Witt (Eds.): Linguistic Modeling of Information and Markup Languages. Dordrecht (Text, Speech and Language Technology Series 40), 203-223.
- Tochtermann, Klaus (1995): Ein Modell für Hypermedia: Beschreibung und integrierte Formalisierung wesentlicher Hypermediakonzepte. Aachen.
- von Hahn, Walther (1983): Fachkommunikation. Entwicklung – Linguistische Konzepte – Betriebliche Beispiele. Berlin. New York (Sammlung Göschen 2223).
- von Hahn, Walther (1997): Das Postulat der Explizitheit für den Fachsprachengebrauch. In: HSK 14.1, 383–389.
- Wiegand, Herbert Ernst (1989): Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. In: HSK 3.1, 409-462.
- Wiegand, Herbert Ernst (1996): Über usuelle und nichtusuella Benennungskontexte in Alltag und Wissenschaft. In: Clemens Knobloch & Burkhard Schaefer (Hrsg.): Nomination – fachsprachlich und gemeinsprachlich. Opladen, 55-103.
- Wiegand, Herbert Ernst (2002): Zur Makrostruktur und zu den äußeren Zugriffsstrukturen im DE GRUYTER WÖRTERBUCH DEUTSCH ALS FREMDSPRACHE. In: Ders. (Hrsg.): Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des „de Gruyter Wörterbuchs Deutsch als Fremdsprache“. Tübingen (Lexicographica. Series Maior 110), 413-442.