

Relationen- und Definitionsanalyse der Kernkorpustexte

Projekt
**Hypertextualisierung auf
textgrammatischer Grundlage**
(www.hytex.info)

Sandra Wellinghoff

2006

Dieses Dokument beschreibt die Ergebnisse einer Pilotstudie zur Extraktion lexikalischer Relationen auf der Basis eines Fachtextkorpus mit annotierten definitiven Textsegmenten.

Die Grundlage für die im Folgenden beschriebene Auswertung bildet ein Korpus mit 20 Fachtexten, in denen 174 vertextete prototypische Definitionen inklusive ihrer Binnenstruktur manuell auf Basis der in Reißwenger/Wellinghoff 2005 beschriebenen Annotationsrichtlinie II annotiert wurden.

Es wurde davon ausgegangen, dass die Struktur einer Definition sowie die als Definitoren verwendeten Ausdrücke Hinweise auf lexikalische Relationen zwischen in der Definition verwendeten lexikalischen Ausdrücken geben können. Für den Ausbau des terminologischen Wortnetzes im *HyTex*-Projekt kann eine Extraktion der entsprechenden Relationen inklusive der beteiligten Relationspartner eine wichtige Ressource darstellen.

So bezeichnet z.B. im folgenden definitiven Textsegment der Definitor *besteht aus* eine Holonymie-Relation:

„Das World Wide Web besteht heute aus Texten, die auf Servern abgelegt und durch Verknüpfungen zu einem weltumspannenden Netz miteinander verbunden sind.“

Ein Definiens-Segment, das nach dem Prinzip „genus proximum – differentia specifica“ aufgebaut ist, enthält stets ein Hyperonym zum im Definiendum genannten terminologischen Ausdruck. In der folgenden Definition ist z.B. *Zeichen* das Hyperonym zum Definiendum *Verknüpfungsanzeiger*. Im Relativsatz wird dieser Oberbegriff, das *genus proximum*, näher differenziert (*differentia specifica*):

„Ein Verknüpfungsanzeiger ist also zunächst immer ein Zeichen, das sich als Ikon oder Symbol identifizieren lässt und darüber hinaus die Funktion und Bedeutung eines Index hat.“

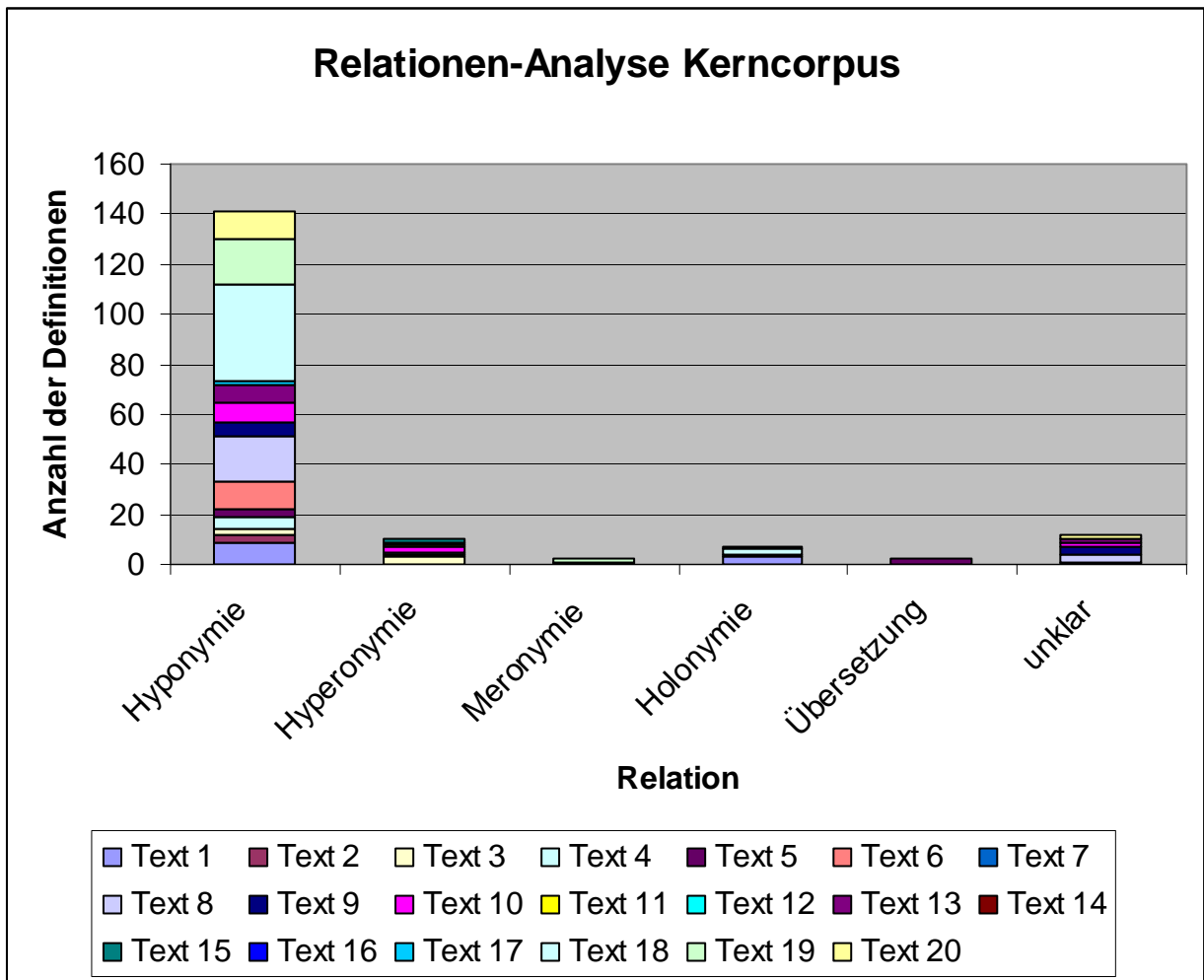
Die manuelle Auswertung der Relationen in den 174 Definitionen der 20 Kernkorpus-texte ergab Folgendes:

Die Hyponymie-Relation ist in 141 Definitionen nachweisbar und damit der am häufigsten innerhalb von Definitionen auftretende Relationstyp. Dies ergibt sich aus der

o.g. Struktur des Definiens (genus proximum/differentia specifica). Alle anderen Relationstypen treten im Verhältnis zur Gesamtzahl der Definitionen nur selten auf:

	Hyponymie	Hyperonymie	Meronymie	Holonymie	Übersetzung	unklar	SUMME
Text 1	9	0	0	3	0	0	12
Text 2	3	0	0	0	0	0	3
Text 3	2	3	0	0	0	0	5
Text 4	5	1	0	0	0	0	6
Text 5	3	0	0	1	2	0	6
Text 6	11	0	0	0	0	1	12
Text 7	0	0	0	0	0	0	0
Text 8	18	1	0	0	0	3	22
Text 9	6	0	0	0	0	3	9
Text 10	8	2	0	0	0	2	12
Text 11	0	1	0	0	0	0	1
Text 12	0	0	0	0	0	0	0
Text 13	7	1	0	0	0	1	9
Text 14	0	0	0	0	0	0	0
Text 15	0	1	0	0	0	0	1
Text 16	0	0	0	0	0	0	0
Text 17	1	0	0	0	0	0	1
Text 18	39	0	1	2	0	0	42
Text 19	18	0	1	1	0	0	20
Text 20	11	0	0	0	0	2	13
SUMME	141	10	2	7	2	12	174

Die folgende Grafik veranschaulicht die Verteilung der Relationen auf die jeweiligen Texte:



Wie die folgende Tabelle zeigt, ist die Auswahl an Definitoren und Separatoren beschränkt:

	Hyponymie	Hyperonymie	Meronymie	Holonymie	Übersetzung	unklar
-	1	0	0	0	0	0
()	3	0	0	0	0	1
,	1	0	0	0	0	0
.	3	0	0	0	0	1
:	5	0	0	0	0	5
bedeuten	2	0	0	0	0	0
begreifen	1	0	0	0	0	0
bekannt / als	1	0	0	0	0	0
benennen	1	0	0	0	0	0
beschreiben	1	1	0	0	0	0
bestehen / aus	0	0	0	7	0	0
bezeichnen / als	13	1	1	0	0	1
charakterisieren / als	1	0	0	0	0	0
definieren / als	1	0	0	0	0	0
gebrauchen	1	0	0	0	0	0
heißen	1	0	0	0	0	2
nennen	7	0	0	0	1	2
ohne Separator	4	1	0	0	0	0
sein	73	2	1	0	1	0
sein (der Oberbegriff für)	0	3	0	0	0	0
spezifizieren / als	4	0	0	0	0	0
sprechen von	1	0	0	0	0	0
Terminus einführen	1	0	0	0	0	0
verstehen / unter	11	2	0	0	0	0
verwenden / als	3	0	0	0	0	0
vorstellen als	1	0	0	0	0	0
	141	10	2	7	2	12

Die Definitoren *sein*, *verstehen unter* und *bezeichnen als* treten vor allem mit Hyponymie-Relationen auf. In den seltenen Hyperonymie-Relationen wird die Beziehung zwischen definiendum und definiens hauptsächlich durch *sein* etabliert. In zwei Definitionen wird explizit ausgedrückt, dass das Definiendum *der Oberbegriff* des Terms im definiens ist. Die Meronymie-Relationen werden durch die Definitoren *sein* und *bezeichnen/als* hergestellt. Im Fall der Holonymie-Relation entsteht die Zuordnung durch den Definitor *bestehen aus*. In den Übersetzungen finden sich die Definitoren *sein* und *nennen*. Unklare Relationen waren vor allem in den Glossar-Definitionen zu finden und wurden durch einen Doppelpunkt etabliert.

In Hinblick auf eine mögliche Automatisierung der Extraktion lexikalischer Relationen aus definatorischen Textsegmenten kann das Ergebnis dieser manuellen Pilotstudie dazu beitragen zu beschreiben, welche lexikalischen Relationen in Definitionen vorzugsweise unter Verwendung welcher Definitoren etabliert werden. Dies kann für die Formulierung von Suchpatterns bei der Automatisierung eine wichtige Ressource darstellen.