

# Texttechnologie in der Universitären Lehre

---

**Andreas Witt** (*andreas.witt@uni-bielefeld.de*)

*Bielefeld University*

**Dieter Metzger** (*dieter.metzger@uni-bielefeld.de*)

*Bielefeld University*

---

## Einleitung

An der Universität Bielefeld (eine deutsche Universität mit ca. 16 000 Studierenden) ist an der Fakultät für Linguistik und Literaturwissenschaft seit wenigen Jahren der Bereich Texttechnologie in der universitären Lehre vertreten. Die Studienmöglichkeiten dieses Faches, die in dieser Form einzigartig sind, sollen auf einem bilingualen Poster (Englisch und Deutsch) erstmals auf einer wissenschaftlichen Tagung vorgestellt werden. Wir sind davon überzeugt, dass unsere Erfahrungen für die Fachkolleginnen und Kollegen an anderen Instituten sehr interessant sind, da es sich um eine innovative curriculare Umsetzung von Inhalten aus dem Bereich Humanities Computing handelt.

Die Texttechnologie an der Universität Bielefeld beschäftigt sich unter anderem mit Eigenschaften von Texten sowie der Modellierung und der Strukturierung textueller Information. Institutionell ist die Texttechnologie dem Arbeitsbereich Computerlinguistik und damit mittelbar dem Fach Linguistik zugeordnet. In der Forschung ist die Texttechnologie seit der Mitte der 1990er Jahre an der Universität Bielefeld vertreten. Zum Wintersemester 1999/2000 begann die Etablierung der Texttechnologie in der Lehre, und zwar in Form eines Magisternebenfachstudiengangs. Dies bedeutete, dass Texttechnologie mit einem geisteswissenschaftlichen Hauptfach und einem weiteren Nebenfach verbunden werden musste. Ein vollständiges Magisterstudium dauert in der Regel 9-10 Semester. Vor ca. zwei Jahren wurden u.a. auf Grund von Internationalisierungsbestrebungen die Magisterstudiengänge in Bielefeld abgeschafft und die Studienmöglichkeiten auf Bachelor- und Masterstudiengänge umgestellt. Dies ist Teil einer generellen Entwicklung in Europa, die 1999 in der Bologna Deklaration von 29 europäischen Ländern fest vereinbart worden ist. Im Zuge dieses Prozesses sollen bis 2010 alle Hochschulstudiengänge auf Bachelor und Master umgestellt worden sein.

Texttechnologie kann seit dem Wintersemester 2002/2003 als eigenes B.A. Nebenfach studiert werden. Dieses Nebenfach kann mit unterschiedlichen B.A.-Kernfächern, z.B. Germanistik, Anglistik, Geschichte, Sozialwissenschaften oder Philosophie kombiniert werden. Die Lehrinhalte des Nebenfachstudiengangs sind modular organisiert.

Nachfolgend sollen diese Module kurz beschrieben werden.

## Module des Studiums

### Grundlagen der Texttechnologie

*Computerpropädeutikum (3 Leistungspunkte)*

*Einführung in die Texttechnologie (4 Leistungspunkte)*

*Textstruktur und Textsatz (4 Leistungspunkte)*

*Hypertext (4 Leistungspunkte)*

Im ersten Studiensemester findet ein Computerpropädeutikum statt, das grundlegende praktische Kenntnisse im Umgang mit Computern vermittelt und dabei auf die Verarbeitung von Textdaten ausgerichtet ist (Betriebs- und Dateisystem, Texteditoren, reguläre Ausdrücke). In der Veranstaltung *Einführung in die Texttechnologie* werden die Eigenschaften elektronischer Texte und die Methoden ihrer Erstellung behandelt. Der Begriff des elektronischen Dokuments, die Trennung von Struktur und Layout, die Zeichencodierung sowie Typographie werden ebenfalls angesprochen. Diese Thematiken werden in der zweistündigen Veranstaltung *Textstruktur und Textsatz* weiter vertieft. Das Seminar *Hypertext* behandelt u.a. Grundlagen dieser Textsorte (von den verschiedenen Begriffsdefinitionen und bis zur Geschichte), konkrete, d. h. praktisch angewendete, Hypertextsysteme und textlinguistische Themen wie z.B. Kohärenz in Hypertexten.

### Formale Methoden der Linguistik

*Formale Methoden I, II, III (je 3 Leistungspunkte)*

Methoden aus der Mathematik, der Informatik und der Logik werden in drei teilweise aufeinander aufbauenden Seminaren besprochen. Zu den Methoden zählen die Mengentheorie als Sprache zur Formulierung von Strukturen, der Umgang mit der Hierarchie der formalen Sprachen und deren Grammatiken (der Chomsky-Hierarchie) und die Verwendung formaler Automaten. Des Weiteren wird in die Aussagen- und Prädikatenlogik eingeführt. Darüber hinaus werden Attribut-Wert-Strukturen und deren Relationen vorgestellt und deren Operationen (vor allem die Unifikationsoperation) eingeübt.

### Programmierung für die Texttechnologie

*Einführung in die Programmierung (12 Leistungspunkte)*

---

Am Beispiel einer konkreten, jeweils aktuellen und für das Humanities Computing relevanten Programmiersprache wird sich in diesem Modul mit den grundlegenden, allgemeinen Prinzipien der Programmierung auseinandergesetzt. Die Vermittlung der programmiersprachlichen Paradigmen steht neben dem Erlernen einer konkreten Programmiersprache (z.B. Java, Perl, PHP oder Prolog) im Zentrum des Moduls. Anwendungsschwerpunkt der Beispiele und der Übungen bildet die Verarbeitung (z.T. strukturiert annotierter) Textdaten. Eine sehr umfangreiche Programmieraufgabe muss *nach Abschluss* des Seminars von Studierenden in kleinen Gruppen gelöst werden.

### **Informationsstrukturierung und Auszeichnungssprachen**

*Auszeichnungssprachen (4 Leistungspunkte)*

*Informationsstrukturierung (8 Leistungspunkte)*

Die Auszeichnungssprachen (engl. *Markup Languages*) zählen zu den grundlegenden Standards der Texttechnologie. In dem Seminar Auszeichnungssprachen wird XML, aber auch SGML, besprochen, wobei die Themen maschinelle Verarbeitung von ausgezeichneten Dokumenten, Schemasprachen, Parser, Transformation und Formatierung von besonderer Relevanz sind.

In dem Seminar *Informationsstrukturierung* werden die dann bekannten Auszeichnungssprachen verwendet, um Informationen, die aus sehr unterschiedlichen Domänen stammen, zu strukturieren. Hierbei wird den Studierenden die Möglichkeit geboten, die Inhalte ihres Kernfachs (bei nahezu allen Studierenden ist dies eine Geisteswissenschaft) als allgemeine Information aufzufassen und zum Gegenstand ihrer Informationsmodellierung werden zu lassen. Die Richtlinien der Text Encoding Initiative werden in diesem Zusammenhang genauer angesehen. In dem Seminar werden Kleingruppen gebildet, die — *kollektiv* — bestimmte Informationsmodellierungsaufgaben lösen und der gesamten Gruppe präsentieren.

### **Wahlpflichtmodule**

Neben den bisher genannten Modulen, die alle studiert werden müssen, ist von den Studierenden noch eines von drei weiteren Modulen zu wählen. Sie können sich entscheiden, ob sie für ca. 3 Semester Sprachkurse (insbesondere Sprachen mit nicht-lateinischen Schriftsystemen, wie Japanisch oder Arabisch) besuchen möchten, ob sie sich mit Empirischen Methoden (u.a. Statistik) oder mit den linguistischen Beschreibungsebenen (z.B. Semantik und Syntax) intensiv beschäftigen möchten.

## **Studieninfrastruktur und weitere Studienmöglichkeiten**

Die Mehrzahl der Seminare enthält sehr praxisnah ausgerichtete Lehrkomponenten. An der Fakultät für Linguistik und Literaturwissenschaft ist in engem Zusammenhang mit dem Aufbau des Studiengangs eine Infrastruktur errichtet worden, die eine derartige praxisnahe Ausbildung ermöglicht. So stehen zwei Multimediaseminarräume (mit 40 und mit 80 Arbeitsplätzen an 20 bzw. 40 Rechnern, einem didaktischen Netz, einem Intranet und selbstverständlich mit Internetanbindung), Rechnerarbeitsplätze für das Selbststudium, sowie Arbeitsplätze im Hochschulrechenzentrum zur Verfügung. Erst durch diese Infrastruktur wurde es möglich, die Praxisnähe auch bei der derzeitigen durchschnittlichen Seminargröße von 60 bis 80 Personen sicherzustellen.

Neben dem B.A.-Nebenfachstudiengang ist es auch möglich, Texttechnologie als so genanntes Profil im B.A. Studiengang Linguistik zu studieren. Darüber hinaus wird an der Universität Bielefeld der M.A.-Studiengang "Interdisziplinäre Medienwissenschaft" angeboten, der fakultätsübergreifend verankert ist. Dieser modular organisierte Masterstudiengang, der auch texttechnologisches Modul beinhaltet, stellt eine von mehreren weiterführenden Studienmöglichkeiten dar, die von den Absolventen des Texttechnologiebachelorstudiengangs gewählt werden können. Es ist natürlich auch möglich nach dem B.A.-Abschluss einen Beruf zu ergreifen. Da sich insbesondere auch die klassischen Berufsfelder für Geisteswissenschaftler/innen (z. B. in den Verlagen oder im Bereich Journalismus) durch den massiven Einzug der Informationstechnologie in den vergangenen Jahren sehr stark verändert haben, sind die Absolventen der Texttechnologie auch für diese (bei den Studierenden der Geisteswissenschaften besonders beliebten Berufe) sehr gut vorbereitet. Es existieren zwar bisher keine Verbleibestatistiken der Studierenden, erste Erfahrungen zeigen aber, dass das Nebenfach Texttechnologie den Studierenden tatsächlich das berufsrelevante Wissen vermittelt, das sich viele Firmen von Absolventen geisteswissenschaftlicher Studiengänge wünschen.