

Prof. Dr. Leslie Siegrist / Sabine Bartsch, MA  
Institut für Sprach- und Literaturwissenschaft  
Technische Hochschule Darmstadt  
D-64289 Darmstadt

## Texplorer Generierung computerunterstützter Lehr- und Lernmaterialien für den fachfremdsprachlichen Deutschunterricht.

Ein deutsch-taiwanesisches Projekt für den fachsprachlichen computerunterstützten  
Fremdsprachenunterricht

Fachbezogener Fremdsprachenunterricht leidet oft unter einem Mangel an geeigneten Lehrmaterialien, da die Rolle, die ein spezieller Fachsprachenunterricht - neben oder im Rahmen eines allgemeinsprachlichen Unterrichts - spielen soll, ungeklärt ist bzw. unterschiedlich eingeschätzt wird (vgl. besonders Fluck/Hoberg/Siegrist 1991, Gläser 1991, Fluck 1992, 1998, Hoberg 1990, 1994, 1998). Das von uns entwickelte System Texplorer versucht dieser besonderen Situation und den Anforderungen von (nicht-muttersprachlichen) Lehrenden und Lernenden an den Umgang mit authentischen Fachtextmaterialien in der Zielsprache durch die Bereitstellung einer Umgebung zur computergestützten Aufbereitung geeigneter Textmaterialien Rechnung zu tragen.

Besondere Bedeutung kommt in diesem Kontext der Erstellung didaktisierter Textbanken und der Erforschung ihrer Einsatzmöglichkeiten zu, denn diese Textbanken können für verschiedenartige Unterrichtsformen Materialien liefern, die in unterschiedliche Fachfremdsprachencurricula im In- und Ausland (z.B. Monteiro1990, Pudszuhn 1994, Ramadan1992, Steinmetz 1996) eingebunden und relativ problemlos für die jeweilige Zielgruppe adaptiert werden können. Eine solche Textbank hat auch dahingehend Vorteile, dass die Texte bei Bedarf aktualisiert, erweitert, verbessert oder ausgewechselt werden können. Aus diesem Grunde wird hier ein Forschungsprojekt<sup>1</sup> vorgestellt, in dem eine didaktisierte Textbank erstellt und zugleich untersucht werden soll, wie mit Hilfe einer

solchen Textbank fachbezogenes fremdsprachliches Lehren und Lernen optimiert werden kann. Dabei kann auf bisherige Forschungsergebnisse der Kooperation zwischen der Technischen Universität Darmstadt, der National Cheng Kung-Universität Tainan, und sowie auf Forschungsergebnisse der GMD (Gesellschaft für Mathematik und Datenverarbeitung) in Darmstadt aufgebaut werden.

Konkretes Ziel des geplanten bilateralen deutsch-taiwanesischen Projektes "Texplorer - Generierung computerunterstützter Lehr- und Lern-Materialien für den fachfremdsprachlichen Deutschunterricht" ist die Entwicklung eines **automatisierten** Systems zur Erstellung von Lehr- und Lernmaterialien für den Fachsprachenunterricht an Universitäten mit dem didaktischen Ziel der Vermittlung primär rezeptiver fachsprachlicher Fähigkeiten, denen im Bereich Deutsch als Wissenschaftssprache heute eine besondere Rolle zukommt (Ammon 1998). Dabei liegt das Hauptaugenmerk zum einen auf der Entwicklung eines Systems, das basierend auf einem Textkorpus von Fachtexten aus der jeweiligen Fachsprachendomäne in der Lage ist, anhand zuvor spezifizierter sprachwissenschaftlicher und didaktischer Parameter der jeweiligen Zielgruppe angepaßte Sprachlehr- / Sprachlerneinheiten automatisch zu generieren und dem Lehrenden / Lernenden in einer als Lehr-/Lerneinheit aufbereiteten Form zu präsentieren.

Da eine Reihe von Grundkomponenten für dieses System schon aus vorherigen Projekten der einzelnen Projektpartner übernommen und integriert werden können, konzentriert sich dieses Projekt auf drei wesentliche Schwerpunktgebiete:

- die Analyse des Korpus mittels des bereits vorhandenen Textanalyseystems zur Generierung einer fachtextspezifischen Grammatik für die Fachtextanalyse;
- die inkrementelle Erweiterung der Korpusrepräsentation anhand didaktischer und linguistischer Parameter basierend auf einer parallel durchgeführten Pilotstudie;
- die Entwicklung eines Lehrer-Lerner-Arbeitsplatzes als Bedienoberfläche mit integriertem multimedialem Wörterbuch.

Basis der Didaktisierung der Textbank und der Entwicklung der Lerneinheiten ist eine projektbegleitende Pilotstudie. Durch exemplarisches Testen der entwickelten Lehrmaterialien im Fachbereich Bauingenieurwesen der National Cheng-Kung Universität in Tainan, Taiwan soll deren Tauglichkeit für den taiwanesischen Unterrichtsalltag evaluiert und Verfahren zur Integration computergestützter Methoden in den fachsprachlichen Fremdsprachenunterricht untersuchend entwickelt werden.

### Aufbau der fachsprachlichen Textbank

Am Institut für Sprach- und Literaturwissenschaft der Technischen Universität Darmstadt wurde vor einigen Jahren begonnen, repräsentative und didaktisierte Textkorpora zu erstellen. Diese Arbeiten wurden von der Volkswagen-Stiftung gefördert und zusammen mit ausländischen Universitäten, insbesondere der Tongji-Universität in Shanghai und der Cheng Kung-Universität in Tainan, durchgeführt.

Es wurden umfangreiche Textkorpora für die Fächer Maschinenbau, Elektrotechnik, Bauingenieurwesen und Wirtschaftswissenschaften zusammen mit den jeweiligen Fachvertretern erstellt. Diese Korpora von mehr als 1000 Texten bzw. Textauszügen, die einen Gesamtumfang von etwa 8 Millionen laufender Wörter ('tokens') umfassen, enthalten alle für das jeweilige Fachgebiet relevanten Textsorten. Anders als bei anderen Textkorpora, wo die Repräsentativität durch die Auswahl der wichtigsten Publikationen auf einem Fachgebiet zustande kommt, wurde die Repräsentativität des Darmstädter Korpus' dadurch sichergestellt, dass Fachvertreter der jeweiligen Fachgebiete Texte und Textauszüge auswählten, die Studierende erfahrungsgemäß während ihres Ingenieurwesenstudiums kennen lernen. Deshalb spiegelt die Textbank auf allen sprachwissenschaftlichen Ebenen der Kern des Fachgebiets für diese Studierenden wider.

Um die Textbank aufbauen zu können, wurden ein Sprachschichtenmodell (z.B. zur Differenzierung zwischen der Sprache des Maschinenbaus im allgemeinen und der Sprache

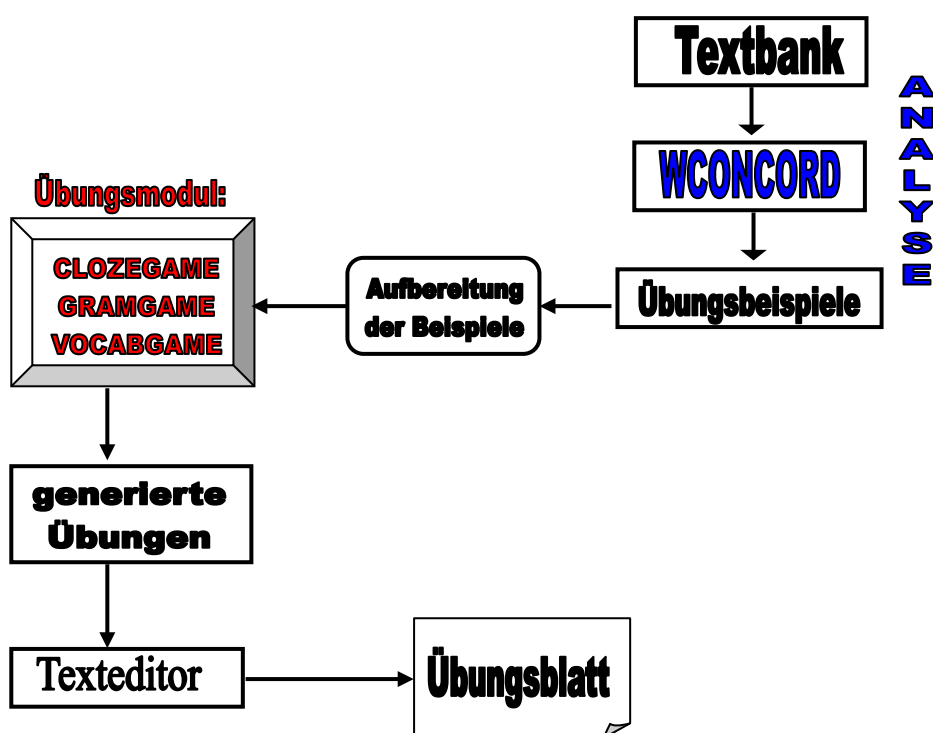
der Fahrzeugtechnik), Textsortenschemata für die genannten Fächer und eine Konzeption für ein Lernerwörterbuch (auf der Grundlage der Textkorpora) erarbeitet.

Der Einsatz authentischer Textkorpora bei der Erstellung von Lehr- und Lernmaterialien für den fremdsprachlichen Fachsprachenunterricht gewinnt angesichts der zunehmenden Verfügbarkeit großer computerbasierter Textkorpora an Bedeutung. Um so verständlicher der Wunsch vieler Lehrender, die den Wert authentischen Textmaterials für die Lernenden und deren natürlichen Gebrauch der Zielsprache erkennen, ihre Lehrmaterialien ebenfalls aus authentischen Textmaterialien zu erstellen und sich dabei verschiedener Übungsformen zu bedienen. Das Lehrende Zeitungstexte aus der Zeitung ausschneiden und in den Unterricht bringen ist kein Novum, und das manche Lehrende Texte sogar abtippen und Lückentexte erstellen ist ebenfalls keine Seltenheit. In den letzten Jahren hat sich im Grenzbereich zwischen computergestützter Sprachwissenschaft und Didaktik ein Forschungszweig entwickelt, der sich mit der Entwicklung von Computerprogrammen für den Unterricht befasst.

### Computerprogramme für Lehrende

In der Literatur zum Einsatz des Computers im Fremdsprachenunterricht findet man selten Informationen über Computerprogramme für Lehrende, mit denen man eigenes Unterrichtsmaterial leicht und effizient erstellen kann. Sogar in der neuesten Literatur werden fast nur Programme für Lernende angesprochen (vgl. Kleinschroth 1996) und vorhandene Forschungsprojekte konzentrieren sich auf fertiggestellte, in sich geschlossene Computerprogramme (computerbasierte Lernsysteme). Der Benutzer kann nur das Material üben, was schon bearbeitet wurde. Neue Beispiele oder Texte können nicht oder nur unter großem Aufwand hinzugefügt werden. Und wenn der Computer als Arbeitshilfe für die Lehrenden erwähnt wird, so meistens nur als Werkzeug, z.B. zur Benutzung von Textverarbeitungsprogrammen, Datenbankprogrammen, Graphikprogrammen,

Sprachanalyseprogramme u.a. (vgl. Rüschoff 1986:79ff.). In Veröffentlichungen von Siegrist (1988, 1994, 1995) werden deshalb Computerprogramme beschrieben, die den Lehrenden bei der Suche und Vorbereitung von geeigneten Unterrichtsmaterialien helfen sollen<sup>2</sup>. Im Rahmen des Textbank-Forschungsprojekts wurden solche Computerprogramme für Fremdsprachenlehrende weiter entwickelt, um eigene Unterrichtsmaterialien schneller und effizienter herstellen zu können. Die Beziehungen zwischen der Textbank und den Programmen kann man graphisch so darstellen:



Zum Status der Systemkomponenten:

Das dargestellte System Texplorer ist ein integriertes modulares computerbasiertes Sprachübungssystem. Die einzelnen in der Grafik rot abgebildeten Systemkomponenten des Übungsmoduls erfüllen einen doppelten Zweck: sie dienen zum einen der Darstellung entsprechender authentischer Sprachmaterialien aus der Textbank in verschiedenen Übungsformen, in dieser Form können Übungen zur Bearbeitung auf konventionellen Übungsblättern (i.e. auf Papier) ausgedruckt werden; zum anderen stellen sie gleichzeitig Teil

einer integrierten On-Line Übungsumgebung für die Übungsbearbeitung und Evaluation am Computerterminal dar.

*Zu den einzelnen Modulen:*

CLOZEGAME ist ein Autoren-Lückentextprogramm, mit dem man eine Vielzahl von Lückenübungen aus Texten der Textbank automatisch generieren kann. Folgende Lückentextarten, die man oft in Lehrwerken findet, können mittels CLOZEGAME generiert werden (der Anfang eines Textes aus der Textbank wird als Beispiel benutzt):

1. Standard Cloze-Texte (bibliographic reference?) (jedes 7. Wort wird ausgelassen)

ÖKONOMISCHE BEDEUTUNG DER ELEKTRISCHEN ENERGIE

Der Energieerhaltungssatz besagt, daß Energie weder gewonnen werden noch verlorengehen kann. Sie kann sich nur \_ (1) einer Energieart in eine andere umwandeln.

\_ (2) einfachen Umwandlungsmöglichkeiten der elektrischen Energie bilden \_ (3) Vorteil dieser Energieart. Die elektrische Energie \_ (4) für jeden Produktionsprozeß notwendig. Sie wird \_ (5) aus der chemischen Energie...

2. C-Test (zweite Hälfte jedes dritten Wortes wird ausgelassen)

Der Energieerhaltungssatz besagt, daß Energie weder gewonnen werden noch verlorengehen kann. Sie kann sich nur v\_\_ (1) einer Energieart i\_ (2) eine andere umwa\_\_\_\_\_ (3).

Die einfachen Umwandlungsm\_\_\_\_\_ (4) der elektrischen Ene\_\_\_\_\_ (5) bilden einen Vor\_\_\_\_\_ (6) dieser Energieart. D\_\_ (7) elektrische Energie i\_\_ (8) für jeden Produkti\_\_\_\_\_ (9) notwendig. Sie wi\_\_ (10) meist aus der chemischen Energie....

3. Wortarten (Präpositionen werden ausgelassen)

Der Energieerhaltungssatz besagt, daß Energie weder gewonnen werden noch verlorengehen kann. Sie kann sich nur \_\_\_ (1) einer Energieart \_\_ (2) eine andere umwandeln.

Die einfachen Umwandlungsmöglichkeiten der elektrischen Energie bilden einen Vorteil dieser Energieart. Die elektrische Energie ist \_\_\_ (3) jeden Produktionsprozeß notwendig. Sie wird meist \_\_\_ (4) der chemischen Energie...

4. Auslassen bestimmter Wörter (Schlüsselwörter)

Der Energieerhaltungssatz besagt, daß Energie weder gewonnen werden noch verlorengehen kann. Sie kann sich nur einer Energieart in eine andere \_\_\_\_\_ (1).

Die einfachen Umwandlungsmöglichkeiten der \_\_\_\_\_ (2) Energie bilden einen \_\_\_\_\_ (3) dieser Energieart. Die elektrische Energie \_ (4) für jeden Produktionsprozeß \_\_\_\_\_ (5). Sie wird meist aus der chemischen Energie...

Lückentexte für Anfänger können auch erstellt werden, in dem man verschiedene Arten von Hinweisen hinzufügt: wenn man entweder die Länge der Lücken angibt oder den ersten Buchstabe hinzufügt, usw., ist der Text entsprechend einfacher zu rekonstruieren. Später dienen die Übungen als Basis für gedruckte Lernmaterialien. Lehrende können nachträglich

mit jedem Textverarbeitungsprogramm Anweisungen, Graphik, usw. zu der generierten Übung hinzufügen.

GRAMGAME (Erzeugen und Darstellen von Grammatikübungen) und VOCABGAME (Erzeugen und Darstellen von Wortschatzübungen) sind zwei weitere Programmmodul. Für GRAMGAME braucht man entweder Einzelsätze oder vollständige Textabschnitte zum Üben grammatischer Themen (Präpositionen, Zeitenbildung, Verbformen, usw.). Die entsprechenden Beispiele werden aus der Textbank extrahiert. Ähnliches gilt für VOCABGAME, ein Programmmodul zur Einübung des fachsprachlichen Wortschatzes.

Im Gegensatz zu CLOZEGAME jedoch kann man die Übungsbeispiele für GRAMGAME und VOCABGAME nicht direkt aus der Textbank holen. Diesen beiden Programmmodulen ist zur Textanalyse das Programm *Concordancer for Windows 3.0* (im folgenden: WCONCORD) vorgeschaltet. Es handelt sich hierbei um ein Konkordanz- bzw. Kollokationsanalyseprogramm, das statistische Informationen über den Text (Worthäufigkeit) bzw. den kontextuellen Gebrauch eines Wortes oder Teil eines Wortes in Form von Konkordanzen, Wortfrequenzliste und Kollokationstabellen mit statistischen Angaben liefert. Lehrende können damit authentische Satzbeispiele aus der Textbank extrahieren und sie in Form von KWIC (*keyword-in-context*) Konkordanzen als Grundlage für die Diskussion / Behandlung lexikalischer oder grammatischer Fragestellungen im Unterricht benutzen (vgl. auch das Konzept des *classroom concordancing* wie es von Tim Johns et al. an der Unit for Overseas Students nach dem Prinzip des *data driven learning* erfolgreich eingesetzt wird; z.B. Talc94 Handbook). Ein durch WConcord aus der Textbank generiertes Beispiel soll dieses Konzept verdeutlichen:

The screenshot shows a window titled 'Concordancer - [Concordance: aus, über, von, mit]'. The main text area contains the sentence: 'Sie kann sich nur [von] einer Energieart in eine andere umwandeln.' Below this is a table with 8 rows, each showing a numbered example sentence with the preposition highlighted in blue. The table columns are: Row Number, Preposition, and Sentence Context.

Row	Preposition	Sentence Context
1.	von	Sie kann sich nur von einer Energieart in eine andere umwandeln.
2.	aus	Sie wird meist aus der chemischen Energie der Brennstoffe (z. B. Braunkohle) oder aus der mechanischen Energie des strömenden Wassers gewonnen.
3.	über	Früher mußte (sic) die Kohle über weite Entfernungen von ihrem Fundort bis zum Betrieb transportiert werden.
4.	von	Früher mußte (sic) die Kohle über weite Entfernungen von ihrem Fundort bis zum Betrieb transportiert werden.
5.	über	Diese wird über Hochspannungsleitungen zu den einzelnen Betrieben übertragen und treibt dort die Maschinen an.
6.	von	Während man bei der Übertragung von mechanischer Energie mit Verlusten bis zu 40% rechnet, liegen die Übertragungsverluste bei elektrischer Energie im Mittel bei 6%.
7.	mit	Während man bei der Übertragung von mechanischer Energie mit Verlusten bis zu 40% rechnet, liegen die Übertragungsverluste bei elektrischer Energie im Mittel bei 6%.
8.	mit	Während man bei der Übertragung von mechanischer Energie mit Verlusten bis zu 40% rechnet, liegen die Übertragungsverluste bei elektrischer Energie im Mittel bei 6%.

WConcord stellt im KWIC-Konkordanzformat die durch den Benutzer gewählten Suchwörter, hier: ‚von‘, ‚aus‘, ‚über‘, ‚mit‘ in einer zentrierten Spalte zusammen mit dem linken und rechten Satzkontext (Horizont) dar. Diese Darstellungsweise verdeutlicht zum einen die syntaktische Struktureinbettung der Suchwörter, erlaubt aber auch die Sortierung nach Wörtern auf Positionen im linken und rechten Kontext des Suchwortes. Auf diese Weise lassen sich bei umfangreicheren Textdatenmengen häufige und typische Verwendungs- und Kontextmuster der Suchwörter identifizieren und die Kookkurrenzhäufigkeit von Suchwort (*node*) und Kontextwörtern / -strukturen statistisch auswerten.

1. Sie kann sich nur von einer Energieart in eine andere umwandeln.
2. Sie wird meist aus der chemischen Energie der Brennstoffe (z.B. Braunkohle) oder aus der mechanischen Energie des strömenden Wassers gewonnen.
3. Früher mußte (sic) die Kohle über weite Entfernungen von ihrem Fundort bis zum Betrieb transportiert werden.
4. Diese wird über Hochspannungsleitungen zu den einzelnen Betrieben übertragen und treibt dort die Maschinen an.
5. Während man bei der Übertragung von mechanischer Energie mit Verlusten bis zu 40% rechnet, liegen die Übertragungsverluste bei elektrischer Energie im Mittel bei 6%.

Beispiele mit Präpositionen müssen nur geringfügig verändert, i.e. markiert, werden, um in GRAMGAME benutzt werden zu können. Computerprogramme wie WConcord vereinfachen die Erstellung von Übungen für den Fremdsprachenunterricht. Partneruniversitäten der Technischen Universität Darmstadt haben in den letzten drei Jahren mit Hilfe der Programme Lehrwerke und Arbeitsmaterial für den fachsprachlichen Unterricht erstellt (Prag, Tallinn, Shanghai, St. Petersburg, Udmurtische Staatsuniversität, Izhevsk). Jedoch mussten oft mehrere Schritte von den Lehrenden manuell ausgeführt werden, um allerlei Arten von Satzbeispielen zu finden. Der Grund dafür liegt im Grundprinzip von WCONCORD, das bei



der Auswahl von Beispielsätzen nur auf uninterpretierte Zeichenketten zugreift und bei komplexeren Zugriffen auf die Textbank nur annäherungsweise passende Beispiele findet. Falls man Passiv-Beispiele für Übungen oder für Klausuren finden möchte, reicht es nicht aus mit dem Konkordanzprogramm WCONCORD nach dem strukturellen Muster *werden, wird* plus *ge\*t* und *ge\*en* zu suchen, da man erstens auch unpassende Beispiele (Futurbeispiele können auch dabei sein) findet und zweitens wichtige passende Beispiele (Verbformen mit Präfixen wie "ver-, zer-", usw.) nicht findet. Unter den gefundenen Beispielen müssen daher aus dem Analyseergebnis von WConcord die passenden Beispielsätze manuell identifiziert werden.

Die Suche nach geeigneten Beispielsätzen wird entsprechend schwieriger, wenn man didaktische und textlinguistische Kriterien für die Auswahl zusätzlich heranziehen möchte. Es wäre z.B. wünschenswert, wenn Lehrende mit einem Computerprogramm automatisch nicht nur Passiv-Beispiele, sondern auch Sätze aus einem bestimmten Subfachgebiet (Feinmechanik als Subfachgebiet der Fachrichtung Maschinenbau) und zweitens geeignete Sätze für Anfänger oder Fortgeschrittene automatisch in der Textbank identifizieren könnten. Voraussetzung für eine automatische Auswahl ist jedoch, dass das Lehrkorpus aufbereitet wird, so dass möglichst umfangreiche Annotationen zu den lexikalischen, morphologischen, syntaktischen, semantischen, textlinguistischen und möglicherweise auch didaktischen Eigenschaften für die Auswahl zur Verfügung stehen. Auf der Basis dieser Annotationen könnte dann ein gezielterer Zugriff auf das Textkorpus erfolgen und dadurch können Lehrende Beispiele aus einer Textbank ohne große Mühe finden.

Für die Erzeugung solcher Annotationen für die deutsche Sprache sind in der Computerlinguistik drei Arten von Programmen<sup>3</sup> mit unterschiedlichem Leistungsumfang entwickelt worden.

Erstens sind *morphologische Analysesysteme* in verschiedenen Varianten verfügbar. Sie ordnen den isolierten Wörtern eines Textes mehrere jeweils mögliche Grundformen und Flexionsmerkmale zu. Die Auflösung von Mehrdeutigkeiten [z.B. "fertigen" → "fertig" (Adjektiv) oder "fertigen" (Verb)] muß dabei meist nach wie vor durch die Lehrenden geschehen. Für größere Korpora kann dieses Verfahren wegen zu vieler mehrdeutiger Fälle in der Praxis unzureichend sein.

Zweitens können *Part-of-speech-Tagger* Wörtern eines Textes unter Beachtung ihres Satzkontextes nach Wortarten taggen. Gute Systeme für das Englische arbeiten mit einer Fehlerquote von unter 5%, inzwischen gibt es auch Adaptionen für das Deutsche (vgl. Feldweg 1995). Nach der Zuordnung der Wortarten bleiben allerdings noch die Konstituentengrenzen und die Funktion der Konstituenten im Satz durch die Lehrenden zu bestimmen (z.B. "Er liest das ganze Buch" - Objekt im Akkusativ vs. "Er liest den ganzen Tag" - Zeitangabe im Akkusativ). Auch dies ist bei größeren Korpora zu aufwendig.

Schließlich haben *Syntax-Analysesysteme*, sog. *Parser* die Funktion, den Wörtern eines Textes ihre Funktion im Rahmen einer Satzstruktur zuzuordnen. Bei dieser Zuordnung werden auch morphologische Mehrdeutigkeiten aufgelöst und Zuordnungen von Wortarten und Konstituenten vorgenommen. Systeme der computerlinguistischen Forschung eignen sich aber wenig für den praktischen Einsatz zur Verarbeitung großer Textmengen mit immer neuen Inhalten wie bei offenen Textkorpora (sog. Monitorcorpora: cf. Sinclair 1991). Oftmals steht die Erforschung bestimmter Formalismen oder Grammatikmodelle im Vordergrund, so dass nur Phänomene, die in dem jeweiligen Rahmen beschrieben sind, verarbeitet werden. Zudem sind viele dieser innerhalb eines bestimmten Forschungsrahmens entwickelten Parser toy-Systeme, die nicht mit großen, strukturell unbegrenzten Sprachdatenmengen umgehen können. Solche Systeme werden zumeist innerhalb eines hochspezialisierten Forschungskontextes entwickelt und sind für den Laien nicht bedienbar. Außerdem führen

Ansätze, in denen auch semantische Phänomene des Satzes (z.B. Zuordnung von Präpositionalphrasen) mit syntaktischen Mitteln beschrieben werden, zu Problemen hinsichtlich der Analyseergebnisse und der Verarbeitungsgeschwindigkeit. Sie erzeugen zu viele und mehrdeutige Lesarten der Sätze. Sofern semantische Kriterien zur Entscheidung eingesetzt werden, wird einerseits die Domäne eingeschränkt, und andererseits werden umfangreiche Lexika mit vorkodierter Information gebraucht. Eine solche Vorgehensweise erscheint zur Erreichung des angestrebten Ziels ungeeignet, so dass nach einem robusteren System Ausschau gehalten werden mußte.

Das *KONTEXT-Textanalysesystem* entwickelt von Haenelt (1994, 1995) ist im Gegensatz zu den im letzten Absatz skizzierten Programmen für den praktischen Einsatz für Fach-Informationstexte entwickelt worden und enthält die drei oben angesprochenen Funktionen. Das System erlaubt eine schnelle und robuste Analyse, die auch für große Textmengen aus unbekanntem Domänen geeignet ist. Dieses Verhalten wird durch eine besondere Analysestrategie und eine effiziente Ressourcen-Verwaltung erreicht. Die syntaktische Analyse beschränkt sich auf die Phänomene, die rein syntaktisch sicher bestimmt werden können. Die syntaktischen und semantischen Beziehungen der Bestandteile werden auf der Basis des semantischen Wortgebrauchs zugeordnet. Dabei wird der Wortgebrauch nach jeder Analyse protokolliert und als Entscheidungshilfe für nachfolgende Analysen genutzt. Dadurch ist das System in der Lage, sich selbst an neue Domänen (Fachgebiete) anzupassen. Das System benötigt jeweils nur einen kleinen Ausschnitt der gesamten Strukturen, um die Bestimmung der syntaktischen Funktionen vornehmen zu können. Eine semantische Wissensbasis wird nicht benötigt. Allerdings können nach Wunsch weitere Lexika angeschlossen werden, um weitere Markierungen zu erzeugen. Ein solches Lexikon ist z.B. die *lexikalische Datenbank TELEX*, in der Lexeme nach der Theorie der semantischen Emphase (Kunze 1991, 1993) modelliert sind.

In einer ersten Erprobungsphase wurde mittels *KONTEXT* eine Auswahl von Texten aus dem Textkorpus der Technischen Universität Darmstadt (verschiedene Fachgebiete, unterschiedliches Fachniveau, verschiedene Textsorten) analysiert. Dabei hat sich herausgestellt, dass das System die Texte robust und schnell (3000 Wörter / Minute) analysiert und die für den Sprachunterricht gewünschten Annotationen bereitstellen kann. Für das Testkorpus wurden jeweils 65% der Texte ohne vorherige Erweiterung der Grammatik annotiert. Das System führt auch partielle Analysen durch und übergeht Teile, für die noch keine hinreichende Modellierung vorliegt. Der Abdeckungsgrad ist inkrementell erweiterbar, indem jeweils Konstruktionen der Fachsprache, die in der Systemgrammatik noch nicht beschrieben sind, zusätzlich beschrieben werden. Diese Erweiterungen beziehen sich auf rein syntaktische Eigenschaften und authentische Textphänomene. Eine Anpassung an semantische Gegebenheiten ist nicht erforderlich. Eine für das Projekt geplante Entwicklung linguistischer Speziallexika (z.B. Klassifizierung von Affixen) eröffnet die Möglichkeit, weitere didaktisch interessante Markierungen vorzunehmen.

Für den Satz "Aus den Zellkernen werden die DNA-Fäden extrahiert." würde *KONTEXT* folgende Annotationen generieren:

#	Wort	Lemma	Wortart	morpho- logische Eigenschaften	Satz- Funktion	Ergänzungen
1	Aus	aus	Präposition		Präposit	PObjekt: 3
2	den	definit	Artikel	casus: dativ numerus: pl genus: masc person: 3	Determ	-
3	Zellkernen	Zellkern	Nomen	casus: dativ numerus: pl genus: masc person: 3	Pobjekt	Determ: 2
4	werden	werden	Hilfsverb	tempus: präs fin: finit person: 3 numerus: pl	Prädikat	AdPräd: 7

				gverbi: passiv		
5	die	definit	Artikel	casus: nomin numerus: pl genus: masc person: 3	Determ	-
6	DNA-Fäden	DNA-Faden	Nomen	casus: nomin numerus: pl genus: masc person:3	Subjekt	Determ: 5
7	extrahiert	extrahieren	Verb	fin: partizip	Adpräd	Subjekt: 6 Präposit: 1
8	.	.	Punkt		.	Prädikat: 4

Diese Annotationen werden zusammen mit den Sätzen aus dem Corpus in einer Datenbank abgelegt. Auf Basis der so generierten annotierten Textdatenbank könnten dann gezielt Sätze mit/ohne Passiv, komplexe Nominale, Infinitivsätze usw. herausgefiltert werden. Ferner könnten auf einer solchen Basis formale Beschreibungen didaktischer Kriterien entwickelt werden (z.B. Sätze für Anfänger → Sätze mit Prädikat, Subjekt und Objekt im Aktiv; Sätze mit bereits im Kurs behandeltem Vokabular usw.). Entsprechende Programmmodule könnten dann als Bausteine zur Handhabung der Textbank programmiert werden.

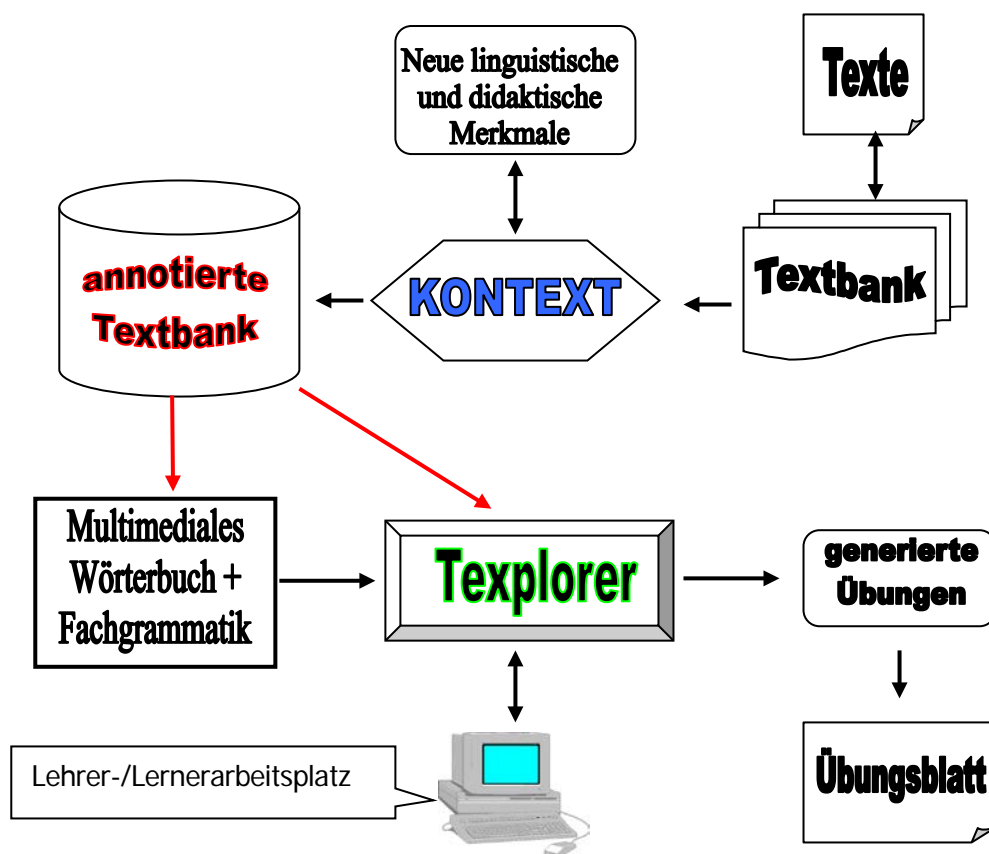
### Automatisierung des Computerprogramms Texplorer

Im geplanten Projekt wird ein computerbasierter Lehrer-Lernerarbeitsplatz zur Integration des didaktisierten Fachsprachenkorpus mit den Analysekomponenten und der daraus resultierenden Möglichkeiten der Aufbereitung und Präsentation der Sprachlehrmaterialien in Form verschiedener Übungstypen in elektronischer Form und in Form gedruckter Arbeitsblätter entwickelt. Der Lehrer- / Lernerarbeitsplatz dient zum einen als Schnittstelle zwischen Lehrenden, Textkorpus und Analysekomponente, so dass es Lehrenden ermöglicht wird ohne Kenntnis einer komplizierten Abfragesprache (*query language*) und ohne direkte manuelle Intervention anhand einer Reihe durch das Programm vorgegebener Optionen und Parameter Vorgaben über die Zielgruppe auszuwählen und entsprechend geeignete

Übungstypen zur Auswahl angeboten zu bekommen. Die entsprechenden Übungen können sodann, in Abhängigkeit von den zur Verfügung stehenden technischen Ressourcen, i.e. Rechnerkapazitäten der Lehrereinrichtung, entweder zur online Bearbeitung direkt am Computer durch die Lernenden bearbeitet werden oder diesen in gedruckter Form zur schriftlichen Bearbeitung vorgelegt werden. Den Lernenden dient der Lehrer- / Lernerarbeitsplatz als Schnittstelle zu den angebotenen Lehrmaterialien, also als Bearbeitungsmedium und als Schnittstelle zu erweiterten online Ressourcen wie zum Beispiel einem multimedialen Wörterbuch. Gleichzeitig kann bei der *online* Bearbeitung der Materialien eine Erfassung der bearbeiteten Materialien zwecks Evaluation der Lernerprogression dienen, was sich besonders bei der erfahrungsgemäß geringen Anzahl der Klassenkontaktstunden im fachsprachlichen Fremdsprachenunterricht im Hinblick auf klassenexterne Selbststudieneinheiten anbietet.

#### Das Textcorpus als Basis für den Lehrer-/Lernerarbeitsplatz

Ausgangspunkt und Basis des Lehrer- / Lernerarbeitsplatzes dessen Grundkonzeption hier erörtert werden soll, ist das wie oben beschrieben annotierte fachsprachliche Textkorpus, das anhand der im Rahmen einer wissenschaftlichen Didaktikkonzeption erarbeiteten Parameter wie Lernerprogressionsstufen, Textklassifikation, Analyse der Korpustexte auf Basis grammatischer und lexikalischer Parameter, annotiert wird, und auf Basis dieser Annotationen von der Analysekomponente durch eine interne Abfragesprache (*query language*) in verschiedene Übungstypen umgesetzt wird. Der Arbeitsplatz sieht dann wie folgt aus:



#### Zusammenspiel der Systemkomponenten:

Im ersten Schritt wird das Textkorpus mit den grundlegenden linguistischen und didaktischen Kategorien mit *KONTEXT* annotiert. Falls *KONTEXT* bestimmte Wörter oder linguistische Strukturen nicht annotieren kann, müssen neue Regeln und Kategorien hinzugefügt werden. Durch dieses inkrementelle Vorgehen ist es möglich, neue Erkenntnisse aus einer von Lehrenden in Taiwan zu durchführende Pilotstudie im Unterricht sukzessiv zu implementieren und auf ihre Tauglichkeit für das zugrundeliegende Lernermodell zu testen und die Materialien auf neue Lernergruppen und deren Bedürfnisse anzupassen. Gleichzeitig erlaubt dieses offene Verfahren die dynamische Erweiterung des Textkorpus, indem in neuen Texten eventuell beobachtete Phänomene und Strukturen ebenfalls in die Annotationen aufgenommen werden können. Damit überwindet das System *TEXPLORER* eine der Schwächen bisheriger Sprach-Lehr-Lernsoftware, die typischerweise gegenüber der Implementierung neuer Phänomene und Anforderungen rigide ist.

Mittels der linguistischen Analysekomponente kann auch eine eigenständige Grammatik des fachsprachlichen Korpus' erstellt. Diese Grammatik erlaubt es, spezielle fachsprachliche Phänomene in Abweichung von der Allgemeinsprache zu identifizieren, da diesen im fachfremdsprachlichen Unterricht besondere Beachtung zukommen muss. Eine wichtige wissenschaftliche Fragestellung wird sich deshalb mit der Erfassung spezifisch fachsprachlicher Phänomene in Textkorpora beschäftigen, um charakteristische Unterschiede zwischen Fach- und Allgemeinsprache und deren Bedeutung im Fachfremdsprachenunterricht zu untersuchen. *KONTEXT* bildet das Korpus mit den aufgenommenen Phänomenen in einer Datenbankstruktur ab (siehe oben), wobei neben den grammatischen und lexikalischen Annotationen auch didaktische Parameter wie Informationen aus der Texttypologie, z.B. Schwierigkeitsgrad der Texte, Eignung für bestimmte Lernergruppen und Lerneinheiten, usw. aufgenommen werden<sup>4</sup>.

Wie erwähnt bildet der Lehrer-/Lerner-Arbeitsplatz ein integriertes System bestehend aus einer Arbeitsoberfläche als Schnittstelle zwischen Lehrenden und dem annotierten Textkorpus einerseits und zwischen Lernenden, Arbeitsmaterialien und Zusatzmaterialien (multimediales Wörterbuch, Grammatikbuch, usw.) andererseits. Diese Schnittstelle liegt in Ansätzen aus Vorarbeiten früherer Projekte vor, im Verlauf des Projektes sollen jedoch die durch die Einzelkomponenten produzierten Übungstypen in ein kohärentes System integriert und in ihrem Erscheinungsbild vereinheitlicht werden. Auch muss die Auswahl der Übungstypen unter didaktischen Gesichtspunkten aufgenommen werden, so dass dem Lehrenden für eine bestimmte Lektion nicht nur die textuell und grammatisch-lexikalisch geeigneten Textmaterialien angeboten werden, sondern auch die entsprechende Präsentationsform in Gestalt von für ein zu übendes Phänomen geeigneten Übungstypen ausgewählt werden.



## Automatische Übungsgenerierung

Eine der grundlegend innovativen Eigenschaften des im Rahmen dieses Projektes angestrebten Systems ist deshalb die **automatische** Generierung von Übungen anhand didaktischer Parameter auf der Basis authentischer Textkorpora aus allgemein- und fachsprachlichen Domänen. Die Erstellung von Lehr-/Lernmaterialien, aber auch die Klassifikation der Texte nach Textsorten und Eignung für verschiedene Lernerprogressionsstufen werden weitestgehend automatisiert, so dass dem Lehrenden einerseits die Möglichkeit gegeben wird mit authentischen Texten bestimmter fachsprachlicher Gebiete auch schon ab einem relativ niedrigen Progressionsniveau zu arbeiten, andererseits aber eine Auswahl frei wählbarer Übungstypen automatisch mit Hilfe des Systems generiert und für Online-Präsentation oder gedruckte Präsentation zusammen mit entsprechenden Zusatzmaterialien (z.B. Multimediawörterbuch und audiovisuelle Unterstützung für die Online-Arbeit wie z.B. Zusatzmaterialien zur Fachlexik und Bildmaterialien, Videosequenzen, usw.) aufbereitet werden kann.

Durch eine so gestaltete Automatisierung bei der Suche nach authentischen und geeigneten Lehr- / Lernmaterial könnten viele Lehrende ihren Unterricht nicht nur den Lernzielen entsprechend effizienter, sondern auch zielgruppenspezifischer und kreativer gestalten.

## Bibliographie

- Ammon, Ulrich. 1998. *Ist Deutsch noch internationale Wissenschaftssprache?* Englisch auch für die Lehre an den deutschsprachigen Hochschulen. Berlin, New York: de Gruyter.
- Fechner, Jürgen (Hrsg.).1994. *Neue Wege im Computergestützten Fremdsprachenunterricht.* Berlin/ München: Langenscheidt.
- Feldweg, Helmut. 1995. Implementation and Evaluation of a German HMM for POS disambiguation. In: *Proceedings of the ACL SIGDAT Workshop*, Dublin 1995.

- Fluck, H.-R..1997. *Fachdeutsch in Naturwissenschaft und Technik*. Aktual. Aufl., Heidelberg: Groos.
- Fluck, H:\_R. 1998: Bedarf, Ziele und Gegenstände fachsprachlicher Ausbildung. In: *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*, hrsg. von Lothar Hoffmann, Hartwig Kalverkämper, H.E. Wiegand, Bd.14.1, Berlin, New York: de Gruyter, 944-954.
- Fluck, Hans-Rüdiger. 1992. *Didaktik der Fachsprachen*. Aufgaben und Arbeitsfelder, Konzepte und Perspektiven im Sprachbereich Deutsch. Tübingen: Narr.
- Fluck, Hans-Rüdiger/Rudolf Hoberg/Leslie Siegrist. 1991. Fachsprachenunterricht für Chinesen. Ein Darmstädter Projekt. In: *Fachsprache*, 13/1-2:48-52.
- Gläser, Rosemarie. 1991. *Fachtextsorten im Englischen*. Tübingen: Narr.
- Haenelt, Karin. 1994. Das Textanalysestystem KONTEXT. Konzeption und Anwendungsmöglichkeiten. In: *Sprache und Datenverarbeitung*, 18/ 1:17-31.
- Haenelt, Karin. 1995. *Das KONTEXT-Modell*. Verarbeitung natürlichsprachiger Texte auf der Basis eines Textmodells. Habilitationsschrift, Universität Heidelberg, März 1995.
- Haenelt, Karin; Könyves-Tóth, Michael. 1991. The Textual Development of Non-Stereotypic Concepts. In: *Proceedings of the 5th Conference of the European Chapter of the Association for Computational Linguistics*, Berlin, April 8-11, S. 263-268.
- Hoberg, Rudolf. 1994. Probleme bei der Erstellung fachsprachlicher Textbanken für Deutsch als Fremdsprache. In: Hoberg, Rudolf (Hrsg.): *Technik in Sprache und Literatur*. Darmstadt: Technische Hochschule Darmstadt, S. 333-348.
- Hoberg, Rudolf.1990. Ist die Behandlung von Fachsprachen im Deutschunterricht für Ausländer wirklich so wichtig? In: Waldemar Pfeiffer (Hrsg.), *Deutsch als Fachsprache in der Deutschlehreraus- und fortbildung*. Poznán.
- Hoberg, Rudolf.1998. Methoden im fachbezogenen Muttersprachenunterricht. In: *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*, hrsg. von Lothar Hoffmann, Hartwig Kalverkämper, H.E. Wiegand, Bd.14.1, Berlin, New York: de Gruyter, 954-960.

- Hoffmann, Lothar/ Kalverkämper, Hartwig/ Wiegand, H.E. (Hrsg.). 1998/99. *Fachsprachen. Languages for special purposes*. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. Berlin, New York: de Gruyter.
- Johns, Tim. 1998. Tim Johns Data-Driven Learning Page. <http://web.bham.ac.uk/johnstf/timconc.htm>. Last updated 18. February 1998.
- Kleinschroth, Robert. 1996. *Neues Lernen mit dem Computer*. Hamburg: Rowohlt.
- Kunze, Jürgen (unter Mitarbeit von Beate Firzlaff). 1993. *Sememstrukturen und Feldstrukturen*. Berlin: Akademie Verlag.
- Kunze, Jürgen. 1991. *Kasusrelationen und Semantische Emphase*. Berlin: Akademie-Verlag.
- Martinek, Zdenek (in collaboration with Leslie Siegrist). 1994. *CLOZEGAME* (1.5). Pilsen: Learsy.
- Martinek, Zdenek (in collaboration with Leslie Siegrist). 1995. *GRAMMATIKGAME* (1.0). Pilsen: Learsy.
- Martinek, Zdenek (in collaboration with Leslie Siegrist). 1997. *WCONCORD* (1.0). Pilsen: Learsy.
- Monteiro, Maria. 1990. *Deutsche Fachsprachen für Studenten im Ausland am Beispiel Brasiliens*. Heidelberg: Groos.
- Pudszuhn, Manfred. 1994. *Fachunterricht versus Sprachunterricht*. Untersuchungen zum studienvorbereitenden Fachsprachenunterricht (DaF) für Ausländer. Frankfurt a.M. u.a.: Lang.
- Ramadan, Hassan. 1992. *Deutsch als Fachsprache in Ägypten*. Begründung eines fachsprachlichen Curriculums (DaF). Heidelberg: Groos.
- Rüschhoff, Bernd, Dieter Wolff, Ernst Ross. 1997. *Fremdsprachenlernen mit dem Computer*. Neue Möglichkeiten zur Förderung von Fremdsprachenkenntnissen für den Beruf. Bielefeld: Bertelsmann.
- Rüschhoff, Bernd. 1986. *Fremdsprachenunterricht mit computergestützten Materialien*. München: Hueber.
- Siegrist, Leslie. 1988. More Vocabulary Programs? in: Udo Jung (ed.). *Computers in Applied Linguistics and Language Teaching*. Frankfurt/Main: Lang, S. 291-296.

- Siegrist, Leslie. 1994. Der Computer und der Fremdsprachenlehrende, in: Jürgen Fechner (Hrsg.). *Neue Wege im computergestützten Fremdsprachenunterricht*. München: Langenscheidt, S. 220-234.
- Siegrist, Leslie. 1995. Computers in Foreign Language Teaching, in: Valentova, Eva/V. Hlavickova (Hrsg.). *Odborný jazyk v podmínkách tržní ekonomiky*. Praha, S. 96-108.
- Siegrist, Leslie (in Zusammenarbeit mit Zdenek Martinek). 1999. *Texplore*. Ein multimediales Computerprogramm für Lehrende. Darmstadt: Technische Universität Darmstadt.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.
- Steinmetz, Maria. 1996. Curriculare Aspekte einer studienbegleitenden Deutschausbildung für chinesische Ingenieurstudenten. In: Fluck, Hans-R. u.a. (Hrsg.), *Deutsch als Fach- und Fremdsprache in Ost- und Zentralasien*. Heidelberg: Groos, 163-196.
- Teaching and Language Corpora (TALC) 94 Handbook. Postscript downloadable from: <http://www.comp.lancs.ac.uk/ucrel/talc94.html>

---

<sup>1</sup> An dem Forschungsprojekt *TEXPLORER* beteiligen sich folgende Wissenschaftler: Prof. Dr. Rudolf Hoberg (TU Darmstadt), Prof. Dr. Hans-Rüdiger Fluck (Universität Bochum), Dr. phil. habil. Karin Haenelt (Universität Heidelberg), Prof. Dr. Jürgen Kunze (Humboldt Universität zu Berlin), Prof. Dr. Jen-Te Chen (National Cheng Kung Universität, Tainan).

<sup>2</sup> Die Computerprogramme *ClozeGame*, *GramGame* und *Wconcord* wurden teilweise in einem früheren Tempus Projekt von Zdenek Martinek (Universität Pilsen) in Zusammenarbeit mit Leslie Siegrist entwickelt.

<sup>3</sup> Die Beschreibung der drei Arten von Programmen und der anschließende annotierte Beispielsatz stammen von Dr. Karin Haenelt.

<sup>4</sup> Die verwendete Analysesoftware *KONTEXT* ist beschrieben in Haenelt (1995) und Haenelt, Könyves-Tóth (1991) und wird im Projektverlauf weiterentwickelt, um auch außersprachlichen, didaktischen und pragmatischen bzw. textlinguistischen Faktoren bei der Textannotation Rechnung zu tragen.