DDD
Deutsch.Diachron.Digital

Das Projekt DDD Deutsch.Diachron.Digital ist die Initiative einer interdisziplinären Forschergruppe, die sich zusammengeschlossen hat, um ein Digitales Referenzkorpus des Deutschen zu entwickeln, das die deutsche Textüberlieferung bis 1900 umfasst. Die Zielsetzung ist eine Digitale Bibliothek, die Faksimiles von Handschriften und Drucken sowie deren Inhalte in Originalsprache und Übersetzung bereitstellt. Für u.a. sprach- und literaturwissenschaftliche Zwecke soll das Material nach neuesten Forschungserkenntnissen aufbereitet und wortweise mit Zusatzinformationen versehen werden.

Ein digitales Korpus für das historische Deutsche
Texte in deutscher Sprache sind seit dem 9. Jahrhundert überliefert. Bislang sind diese Zeugnisse überwiegend in gedruckten Editionen verfügbar. Im Vergleich zu herkömmlichen Korpora bietet das digitale Korpus für die wissenschaftlicheArbeit viele Vorteile:

- ständige Verfügbarkeit im Internet
- optimale Suchmöglichkeiten
- linguistische und literaturwissenschaftliche Informationen
- individuelle Anpassung der Optionen an den jeweiligen Nutzer

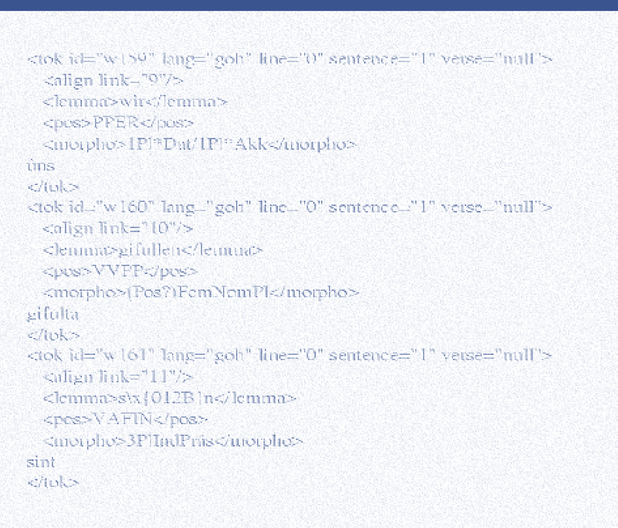
Die Digitalisierung der deutschen Textüberlieferung als kulturelle Aufgabe
Das deutsche Schriftgut vom Mittelalter bis zum 19. Jahrhundert ist uns in Handschriften und Drucken überliefert, deren Schriftzüge zum Teil nur noch Fachleuten vertraut sind. Die Sprachstufen dieser Texte sind die Vorgänger unseres heutigen Deutschen und gewähren wertvolle Einblicke in die Entwicklung der deutschen Sprache. Die Texte sind zugleich Zeugnisse deutscher und europäischer Geschichte und damit ein wesentlicher Bestandteil unseres kulturellen Erbes. Durch die Bereitstellung der Originaltexte soll ein vielseitig nutzbares Forschungswerkzeug entstehen, das den Fragestellungen verschiedenster Wissenschaftszweige genügt.

Die Digitalisierung deutscher Textüberlieferung als Forschungsressource für textbezogene Wissenschaften
Elektronische Texte werden zusammen mit Zusatzinformationen in einer Datenbank gespeichert und über eine Benutzeroberfläche im WWW zugänglich gemacht. Die Struktur der Datenbank erlaubt die Zuweisung äußerst detaillierter Informationen. Diese betreffen sowohl den gesamten Text als auch die Einheiten unterhalb der Textebene (Kapitel, Sätze, Wörter).

Beispiel für Aufbereitung: Hildebrandslied
Alle Texte werden auf mehreren Ebenen mit zusätzlichen Informationen versehen, d.h. die einzelnen Zeichen sind mit den entsprechenden Stellen in der Handschrift verknüpft, wobei bestimmte vorher festgelegte graphische Eigenschaften aus der Handschrift (Farbe, Majuskeln, Rasuren etc.) angegeben werden. Außerdem werden den Wörtern ihre linguistischen Eigenschaften zugeordnet. Sprach- und literaturhistorisch relevante Metainformationen zum jeweiligen Text sind in einem Dateivorspann verzeichnet.

Abbildung 1 stellt eine vereinfachte Tabelle für den ersten Satz aus dem Hildebrandslied dar.

Alle Informationen können innerhalb eines Textes, aber auch zwischen verschiedenen Texten zueinander in Beziehung gesetzt werden. Möchte man z.B. die zeitliche Entwicklung eines Wortes untersuchen, kann man nach der Nennform (z.B. Imbiss) suchen und findet alle Textstellen, in denen das Substantiv in irgendeiner (flektierten) Form vorkommt. So kann beispielsweise die etymologische und semantische Entwicklung des Wortes nachvollzogen werden. Dies veranschaulicht die Bedeutungsveränderung des Wortes „Imbiss" vom Althochdeutschen bis zum Neuhochdeutschen in Abbildung 2.

HU Berlin | Institut für deutsche Sprache und Linguistik | Sprachwissenschaft
Prof. Anke Lüdeling | Frank Henrik Müller | 2093-9799
ddd@hu-berlin.de | http://www.deutschdiachrondigital.de

Multimedia-Förderprogramm 2005