

[erscheint in: Arbeitsbericht 2003 des Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften, im Erscheinen]

Internetbasiertes Artikelredaktionssystem für die Produktion und Publikation von Wörterbüchern in dezentralen Arbeitsstellen

DFG-Projekt der Trierer Arbeitsstelle des Neuen Mittelhochdeutschen Wörterbuchs in Verbindung mit dem Kompetenzzentrum (vgl. Arbeitsbericht 2002, S. 18f.)

Ziel des Projekts ist die Entwicklung eines datenbankbasierten, internetfähigen Artikelredaktionssystems für dezentral arbeitende Wörterbucharbeitsstellen. Das Redaktionssystem soll alle Phasen der Wörterbucharbeit unterstützen: Die Erhebung und Lemmatisierung von Belegen zu einem Stichwort, die schrittweise Strukturierung, Bearbeitung und Kommentierung der Belege zu einem ausgearbeiteten Artikel und schließlich die Publikation des Wörterbuchs in gedruckter und elektronischer Form und des Belegarchivs im WWW.

Das Projekt wird durchgeführt in enger Zusammenarbeit mit den beiden Akademiearbeitsstellen des Mittelhochdeutschen Wörterbuchs in Göttingen und Trier, nach deren Anforderungsprofil und auf deren Datenbasis das Redaktionssystem konzipiert ist, bevor es in einer folgenden Projektphase (3. Jahr) für andere vergleichbare Wörterbuchvorhaben geöffnet und adaptiert werden soll. In den beiden Arbeitsstellen wird im Juni 2004 die erste Doppellieferung des Mittelhochdeutschen Wörterbuchs mit dem Redaktionssystem fertiggestellt. Das Projekt wird seit März 2002 durch die DFG gefördert, nachdem es zuvor mit Mitteln des Forschungsfonds der Universität Trier vorbereitet worden war. Ein Antrag auf Förderung für das dritte Projektjahr wurde bei der DFG eingereicht.

Das Redaktionssystem besteht aus Benutzersicht aus den Kernkomponenten Text- und Belegarchiv, Artikelstrecke, Quellenverzeichnis und Artikeleditor. Das *Text- und Belegarchiv* unterstützt die Lexikographen bei der Erhebung von (lemmatisiertem) Belegmaterial zu einem Stichwort. Dabei ist trunkeierte Suche und Einschränkung der Suche auf frei bestimmbare Teilcorpora möglich. Die präsentierte Ergebnisliste der Belege kann nach verschiedenen Kriterien geordnet werden, z.B. nach Text, Wortform oder Datierung der Quellen. Die Wahl der Länge des Kontextausschnitts kann frei bestimmt werden, zusätzlich ist der Sprung in den elektronischen Volltext möglich, um Belege in ihrem Textzusammenhang beurteilen zu können. Außerdem können Belege von hieraus in den eigentlichen Artikeleditor zur Verarbeitung in einen Wörterbuchartikel kopiert werden. Die *Artikelstrecke* gewährt einen Überblick über die verschiedenen Bearbeitungs- und Korrekturzustände von Artikeln und Stichwortstrecken; sie bildet damit die zentrale administrative Komponente des Systems. Hier können die Lexikographen überblicken, welche Artikel bereits geschrieben wurden, in welchem Bearbeitungszustand sie vorliegen, welche Lemmata (noch) nicht bearbeitet wurden und welche als Verweise angesetzt wurden. Das *Quellenverzeichnis* stellt die Verwaltungs- und Informationskomponente über das Siglenverzeichnis des Wörterbuchs dar. Es enthält

alle nötigen Informationen über die Quellen wie die bibliographische Angabe, Symptomwertklassifikationen, Informationen über die Auswertung der Quellen in anderen Wörterbüchern, die Sigle usw. Ferner unterstützt diese Komponente die aufwendige Pflege und Konsistenzsicherung des Siglensystems. Der *Artikeleditor* ist das Kernstück des System, in ihm können Belege geordnet, bearbeitet und gemeinsam mit lexikographischem Kommentar zu Artikeln verarbeitet werden; gleichzeitig sind sie jedoch gegen nicht regelhafte Bearbeitung geschützt, so daß aufwendige Korrekturgänge der ausgearbeiteten Artikel entfallen. Wie auch im Belegarchiv ist jederzeit die Möglichkeit gegeben, den Belegkontext zu verkürzen, zu verlängern oder den Beleg im Volltext anzuzeigen. Es stehen ferner Funktionen zur Gliederung des Artikels mit automatischer Numerierung bereit. Der Artikeleditor ist tagbasiert konzipiert, so daß alle nötigen Informationspositionen der Artikel durch XML-Elemente ausgezeichnet sind. Die meisten dieser Tags werden vom Artikeleditor automatisch eingesetzt nach Maßgabe derjenigen Funktionen, die der Artikelredakteur aufruft. Diese Tags bleiben im Artikeleditor der Übersichtlichkeit halber unsichtbar. Eine kleinere Gruppe von Elementen der Textauszeichnung kann von den Redakteuren frei (z.B. über Tastatur-Makros) eingegeben werden. Die inhaltliche Auszeichnung der Artikel wird für den Satz der Artikel mit dem TUSTEP-Satzprogramm in typographische Auszeichnung umgesetzt; für eine elektronische Publikation dient sie der Vorbereitung der elektronischen Nutzungsmöglichkeiten wie z.B. der Suche in bestimmten Feldern oder Bereichen der Artikel.

Technisch ist das System als Client/Server-Architektur konzipiert, in dessen Zentrum eine relationale internetfähige Datenbank steht, in der alle Wörterbuchdaten (Artikel, Quellentexte, Stichwortliste usw.) gespeichert sind. Auf den einzelnen lexikographischen Arbeitsplätzen ist die in Tcl/Tk implementierte Clientsoftware installiert, die die oben skizzierten lexikographischen Funktionen zur Verfügung stellt und die Kommunikation mit der zentralen Datenbank regelt. Diese Architektur erlaubt es räumlich getrennten Arbeitsstellen, simultan mit demselben Datenmaterial zu arbeiten. Eine XML-Import- und Export-Schnittstelle gewährleistet die Möglichkeit zur Übernahme von entsprechend kodierten Materialien (Artikel, Texte usw.) in das System und außerdem die langfristige, plattformunabhängige Datenhaltung der Wörterbuchdaten.

Wie im Arbeitsprogramm vorgesehen konnten die zentralen Komponenten des Systems im Jahr 2003 weiter ausgebaut und perfektioniert werden. Dies betrifft vor allem den Artikeleditor und das Quellenverzeichnis. Dabei konnten zahlreiche Verbesserungsvorschläge umgesetzt werden, die sich im täglichen Einsatz des Systems in der Trierer Wörterbucharbeitsstelle ergeben hatten. Die enge zeitliche wie inhaltliche Verzahnung von Entwicklung und tatsächlichem Einsatz des Systems in der Produktion von Artikeln unter Echtbedingungen garantiert einen starken Praxisbezug und einen hohen Qualitätsstandard des Systems. Außerdem wurde die Entwicklung einiger größerer Funktionen in Angriff genommen wie z.B. das Lemmatisierungswerkzeug, das die Zuordnung von Belegen zu einem Stichwort interaktiv unterstützt und die XML-konforme Import- und Export-Schnittstelle für die Wörterbuchdaten.

Das Projekt wurde 2003 auf zwei Tagungen vorgestellt: auf der ACH/ALLC Conference, die vom 29.05. – 02.06. an der University of Georgia in Athens (USA) stattgefunden hat,

in einem Referat von Frank Queens und Ute Recker-Hamm mit dem Titel ‚An Internet-based working environment for the production of dictionaries in distributed locations‘ und im Rahmen einer Präsentation auf der 4. Arbeitstagung deutschsprachiger Akademiewörterbücher in Göttingen vom 9.-12. September.

Homepage des Projekts: <http://www.mhdwb.uni-trier.de/TARes>

[Abbildung folgt]

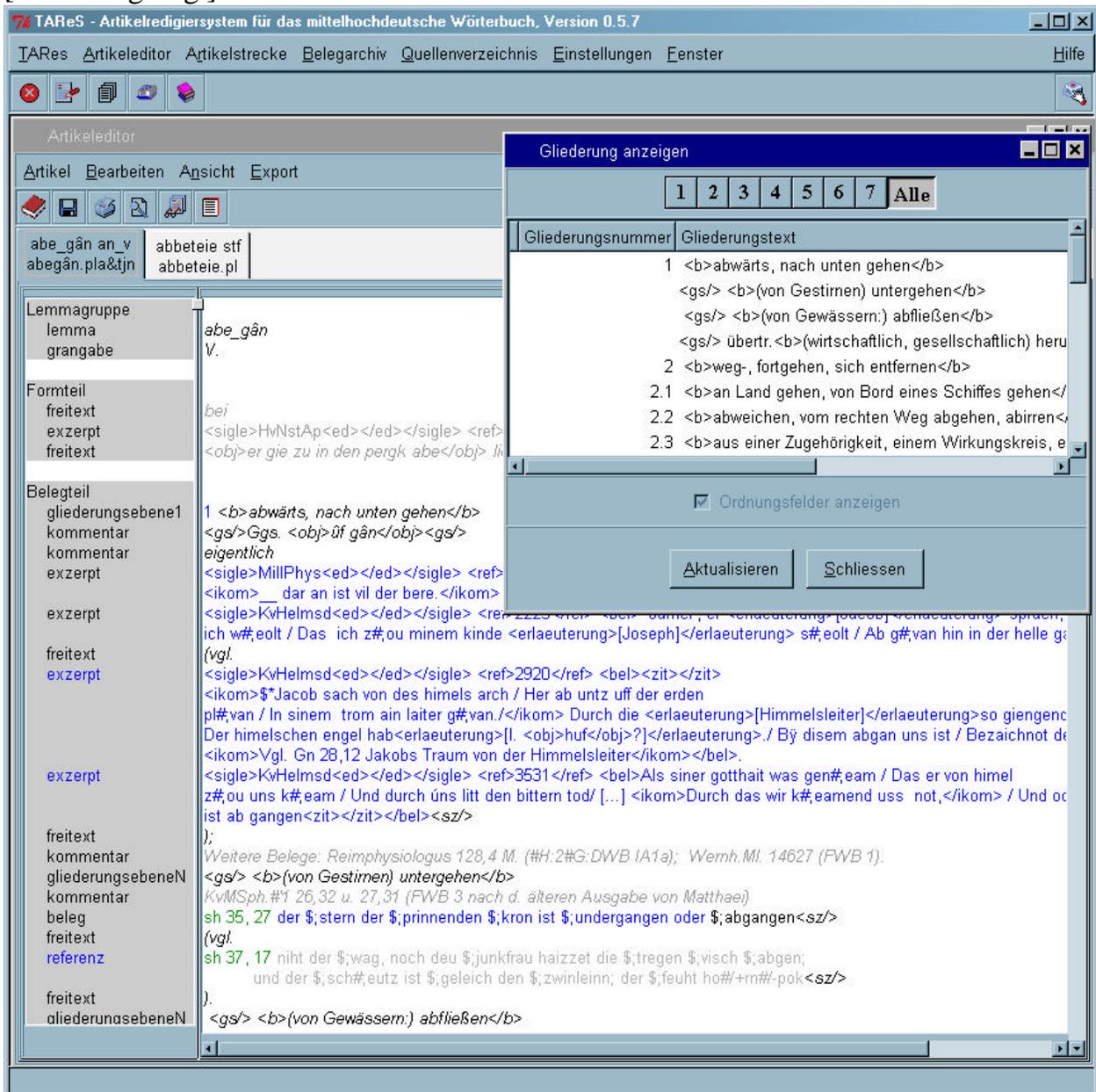


Abb.: Artikeleditor mit Gliederungsansicht