

## **Retrodigitalisierung von Wörterbüchern mit diskursiver Mikrostruktur: Lexikografische und technische Aspekte am Beispiel eines Lehnwörterbuchs**

Peter Meyer, Institut für Deutsche Sprache, Mannheim

Vor allem in älteren Wörterbüchern mit philologischer Ausrichtung ist die Mikrostruktur der Artikel häufig diskursiv und unsystematisch. Eine automatisierte Digitalisierung solcher Wörterbücher mit dem Ziel, ihre logische Struktur zu kodieren, ist nicht möglich; in vielen Fällen ist schon ein Parser für ein manuell nachzubearbeitendes Rohdigitalisat kein realistisches Ziel, weil die Angabetypen des Wörterbuchs nicht klar voneinander abgrenzbar und in den Einzelartikeln nicht eindeutig identifizierbar sind. In solchen Fällen wirft auch eine nachträgliche manuelle Formalisierung der Mikrostruktur große lexikografische Probleme auf.

Für komplexere Anwendungsszenarien wie etwa Abfragen in Webanwendungen kann es dennoch unumgänglich sein, wenigstens sämtliche relevanten in den Artikeln diskutierten Wortformen mit grundsätzlichen diastematischen und morphologischen Informationen sowie ihren Relationen zueinander in einem maschinell lesbaren Format strukturiert zu repräsentieren, etwa durch datenzentrierte XML-Dokumente.

Der Vortrag versucht, die lexikografischen und technischen Möglichkeiten und Grenzen einer solchen teilweisen und manuellen Retrodigitalisierung am Beispiel von Erfahrungen mit einem älteren Wörterbuch deutscher Lehnwörter im Slovenischen (Striedter-Temps 1963) auszuloten. Das Wörterbuch soll in ein Portal von Lehnwörterbüchern mit Deutsch als gemeinsamer Gebersprache eingebunden werden. Die Einzelartikel werden dem Benutzer als Bilddigitalisate zur Verfügung gestellt; die zusätzliche textuelle Retrodigitalisierung ist jedoch für komplexere, insbesondere auch für wörterbuchübergreifende und portalweite, Suchabfragen erforderlich.