

## **Sprache - sprâhe - spriche – Sproch... Die Konzeption einer Metalemmaliste zur Erschließung von Varietäten und Varianz**

### **Abstract**

Im Rahmen des vom BMBF geförderten Verbundprojekts „Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprache und Genomen“ entsteht eine Metalemmaliste, die zum einen als Anknüpfungspunkt für eine vergleichende Untersuchung von Varianz und zum anderen als Verknüpfungswerkzeug sprachwissenschaftlicher Daten dienen soll.

Synchrone und diachrone Varianzen können als Regelzustand einer Sprache angesehen werden. Mit Hilfe einer Metalemmaliste kann diese Varianz erschließbar gemacht und ein dynamischer Zugang erreicht werden. Als Datenbasis hierfür dienen historische Wörterbücher und Idiotika, die am Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften in digitaler Form vorliegen, sowie eine Basislemmaliste des Gegenwartssprache, die am Institut für Deutsche Sprache aus Korpusdaten ermittelt wird.

Varietätenlemmata desselben Etymons (*Sprache, sprâhe, spriche, Sproch*) werden über ein neuhochdeutsches Metalemma (*Sprache*) aufeinander abgebildet.

Es werden verschiedene Verfahren vorgestellt, die diese Zuordnungen generieren.

Diese basieren einerseits auf linguistischem Expertenwissen, andererseits auf maschinell erlerntem, aus den Daten extrahiertem Wissen. Insbesondere wird hier auf Methoden aus der Bioinformatik zurückgegriffen.

Neben dem semasiologisch angelegten Grundgerüst der Metalemmaliste können auch onomasiologische Vernetzungen über Wortkollokationen realisiert werden. Auf diese Art entsteht ein nützliches Tool, das auch als Schnittstelle für weitere Informationssysteme fungieren kann.