

Stefan Engelberg
engelberg@ids-mannheim.de

&

Peter Meyer
meyer-peter@gmx.de

IDS Mannheim

Graphenbasierte ontologische Suchen im *Lehnwortportal Deutsch*

Die Rechercheoptionen im mehrsprachigen *Lehnwortportal Deutsch* (lwp.ids-mannheim.de) sollten auch eine onomasiologische Erschließung des Datenbestands ermöglichen. Diese würde es erlauben, durch Zeiten hindurch und über Sprachen hinweg die Entwicklungen in bestimmten semantischen Entlehnungsdomänen zu verfolgen.

Basiert man eine solche onomasiologische Suche auf einer expliziten, im weitesten Sinne ‚ontologischen‘ Klassifikation der Lexeme, sind verschiedene Entscheidungen zu treffen, u.a.: (i) eigene vs. etablierte Ontologie; (ii) Ein- vs. Mehrebenenontologie; (iii) exhaustive vs. partielle Annotation. In Rahmen des von uns konzipierten Lehnwortportals plädieren wir für eine eigene, partielle Zweiebenenontologie. Annotiert werden soll auf einer thematischen und einer taxonomischen Ebene. Diese in Ontologien nicht immer deutlich durchgeführte Trennung ist aus Gründen semantischer Klarheit unabdingbar und wird z.B. in (Lorentzen & Trap-Jensen 2011) in Hinblick auf onomasiologische Suchen in *Den Danske Ordbog* diskutiert. Mit der im Rahmen der *World Loanword Database* entwickelten Ontologie und den für verschiedene Sprachen verfügbaren WordNets liegen Ontologien beiden Typs vor. Beide wurden im Rahmen des Lehnwortportalprojekts stichprobenbasiert getestet und für wenig geeignet befunden. Das hat vor allem mit dem oft sehr spezifischen Charakter des Lehnwortschatzes zu tun. Lexeme können so zum Teil gar nicht an vorgegebene Kategorien angebunden werden, oder das Raster ist so grob (v.a. in der WOLD-Ontologie), dass alle interessanten semantischen Klassenbildungen verlorengehen. In dem Vortrag werden wir über erste Versuche berichten, ein eigenes Annotationssystem zu entwickeln, das zu einem großen Teil induktiv aus den Spezifika der Lehnwortschätze gewonnen wird, und wir werden die dabei auftretenden Schwierigkeiten darstellen.

Außerdem werden wir die Integration der Annotationen in das Datenmodellierungskonzept des Lehnwortportals skizzieren, das das lexikographische Material des Portals als wörterbuchübergreifendes Netzwerk (gerichtete Graphen) von binären Relationen (Graphenkanten) zwischen gebuchten Wörtern (Graphenknoten; insbesondere Etyma und Lehnwörter mit ihren Varianten, Derivaten usw.) repräsentiert und durchsuchbar macht. Die geplante Integration umfasst vor allem zwei Punkte: (a) Die Einzelbedeutungen (Lesarten) werden nicht mehr wie bisher als Attribute von Knoten dargestellt, sondern sind selber Knoten im Graphen und haben ihrerseits ontologische Annotationen als Attribute. (b) Diese ontologischen Attribute werden als neue Suchoptionen in die erweiterten Suchfunktionen des Portals eingebunden und können dann auch mit den bereits bestehenden Suchmöglichkeiten nach Stichwörtern in den lexikographischen Bedeutungsdefinitionen kombiniert werden. Insbesondere mit der graphenbasierten Abfragesprache des Portals (Meyer 2013) können so beliebig komplexe und granular auf der Ebene von Lesarten von Einzelwörtern definierte Konstellationen semantischen Wandels auch über Entlehnungsketten hinweg in der Datenbasis des Portals gesucht und gefunden werden.

Literatur

Lorentzen, Henrik & Lars Trap-Jensen (2011): There And Back Again – from Dictionary to Wordnet to Thesaurus and Vice Versa: How to Use and Reuse Dictionary Data in a Conceptual Dictionary. In: *Proceedings of eLex 2011*, 175-179.

Meyer, Peter (2013): Advanced graph-based searches in an Internet dictionary portal. In: *Proceedings of eLex 2013*, 488-502.