

Entwicklung eines graphentheoretischen Analysetools zur Überarbeitung des terminologischen Wörterbuchs in Grammis

Grammis ist ein umfassendes grammatisches Online-Informationssystem, das am Institut für Deutsche Sprache, Mannheim, entwickelt wurde und derzeit evaluiert und überarbeitet wird (vgl. Suchowolec et al. (angenommen)). In diesem Paper beschreiben wir die Entwicklung eines graphentheoretischen Analysetools zur Unterstützung dieser Überarbeitung. Graphentheoretische Ansätze finden bereits in funktional ähnlichen Kontexten Anwendung, um Strukturen in komplexen Wissenssystemen aufzudecken (vgl. Mehler 2008, Mehler 2017, Watts und Strogatz 1998). Ein solches Tool wird erstmals im Kontext der linguistischen Terminologie eingesetzt.

Grammis besteht aus mehreren Komponenten, die sich mit unterschiedlichen Aspekten der deutschen Grammatik beschäftigen. Zwei dieser Komponenten sind das grammatische Fachwörterbuch sowie der Thesaurus. Der Thesaurus bildet die Vernetzung der Konzepte untereinander ab, sowohl durch hierarchische Relationen, als auch durch Assoziationsrelationen. Dies wird als Makrostruktur bezeichnet. Das grammatische Fachwörterbuch hingegen besteht aus wissenschaftlichen Volltextartikeln, die einzelne Konzepte für den Nutzer von Grammis näher beschreiben, und wird Mikrostruktur genannt. Die Artikel liegen als Hypertext vor, die durch Verlinkung (Hyperlinks) untereinander verknüpft sind. Die Verlinkung der Konzepte in der Mikrostruktur wird von den Autoren händisch vorgenommen. Da die Links der Mikrostruktur nicht auf dem Thesaurus aufbauen, ist nicht zwingend sichergestellt, dass diese Verlinkung der Makrostruktur entspricht. Der Thesaurus hingegen zeigt die makrostrukturelle Verknüpfung der Konzepte, wie sie in der Datenbank hinterlegt ist. Die angestrebte Überarbeitung dient unter anderem dazu, Makro- und Mikrostruktur in einer gemeinsamen Datenbasis abzulegen und jedem Konzept seine mikrostrukturelle Beschreibung direkt zuzuordnen.

Das Analysetool soll die Verlinkungen der Mikrostruktur und die Relationen der Makrostruktur als Graph darstellen und dadurch multidimensionale Auswertungen der Ressourcen ermöglichen. Eine direkte Gegenüberstellung der Makro- und Mikrostruktur eines gewählten Konzeptes ermöglicht es den Autoren, die Beziehungen dieses Konzeptes zu seinen Nachbarn (sog. Lokale Gruppe) auf einen Blick zu erfassen. Die Visualisierung wird in Form eines gerichteten Graphen realisiert. Insgesamt ergeben sich vier Vergleichsmöglichkeiten: auf horizontaler Ebene, also Makrostruktur des alten Grammis (Ma_{alt}) mit der Makrostruktur der überarbeiteten Einheiten (Ma_{neu}) sowie der Vergleich der jeweiligen Mikrostrukturen (Mi_{alt} mit Mi_{neu}). Außerdem lässt sich auf vertikaler Ebene die Makrostruktur mit der Mikrostruktur vergleichen. Also Ma_{alt} mit Mi_{alt} sowie Ma_{neu} mit Mi_{neu} . Zum Vergleich der Strukturen werden bestimmte Kennzahlen, wie lokaler Clusterkoeffizient, mittlere Weglänge etc. berechnet (vgl. Watts und Strogatz 1998, Newman 2003). Änderungen an den jeweiligen Konzepten werden in Echtzeit dargestellt.

Ein Anwendungsbeispiel ist, dass sich durch den Vergleich von Mi_{alt} und Ma_{alt} (bzw. Mi_{neu} und Ma_{neu}) für den Autor nachvollziehen lässt, inwiefern der verfasste Text der makrostrukturellen Vernetzung entspricht. Dies ist unter anderem deshalb wichtig, da es ein Kriterium ist, dass Benutzer auf der Mikrostruktur möglichst direkt zu den jeweiligen Hyperonymen, Kohyponymen, Hyponymen und Meronymen navigieren können. Dies kann anhand des lokalen Clusterkoeffizienten des jeweiligen Hyperonyms operationalisiert werden. Eine Gegenüberstellung der lokalen Clusterkoeffizienten eines Konzeptes in Mi_{alt}

und Mineu zeigt zudem, ob die Überarbeitung in dieser Hinsicht eine Verbesserung erbracht hat. Darüber hinaus bieten sich weitere Möglichkeiten zur Anwendung der Graphenanalyse in der Überarbeitung der Ressourcen, beispielsweise eine Überprüfung, welche in der Mikrostruktur gesetzten Verlinkungen auf ein Konzept verweisen, für das bereits ein ausformulierter Erklärungstext existiert.

Literatur

- Mehler, Alexander (2008). »Structural similarities of complex networks: A computational model by example of wiki graphs«. In: Applied Artificial Intelligence 22.7-8, S. 619–683.
- (2017). »Zur Struktur und Dynamik von kollaborativ erzeugten Lexikalischen Netzwerken«. In: Vortrag gehalten auf 53. Jahrestagung des Instituts für Deutsche Sprache (IDS), Mannheim, 14.3.–16.3.2017.
- Newman, M. E. J. (2003). »The Structure and Function of Complex Networks«. In: SIAM Review 45.2, S. 167–256.
- Suchowolec, Karolina, Christian Lang, Roman Schneider und Horst Schwinn (angenommen). »Complexity from Text to Data Model. Adding Machine-Oriented Features to a Human-Oriented Terminology Resource«. In: Proceedings of the 1st Conference on Language, Data and Knowledge. Galway, Irland 19.06.—20.06.2017. Jorge Gracia, Francis Bond, John P. McCrae, Paul Buitelaar, Christian Chiarcos und Sebastian Hellmann (Hrsg.). Lecture Notes in Computer Science: Springer.
- Watts, Duncan J. und Steven H. Strogatz (1998). »Collective dynamics of ‘small-world’ networks«. In: Nature 393.6684, S. 440–442.