

Kausalitätsmarker als Kohärenzmittel und ihre Formalisierung für die automatische Textanalyse

Kurztitel: „kausale Kohärenz“

Gemeinschaftsprojekt (DFG gefördert)

Prof. Dr. Manfred Stede
Universität Potsdam, Institut für Linguistik
stede@ling.uni-potsdam.de

Projektgruppe Handbuch der deutschen Konnektoren
Institut für Deutsche Sprache, Mannheim
E-Mail: breindl@ids-mannheim.de

Projektbeginn: Potsdam 1.10.2006
Mannheim 1.1.2007

Zusammenfassung

Verfahren der automatischen Textanalyse haben sich in den letzten 15 Jahren vorwiegend mit statistischen und selbstlernenden Methoden beschäftigt, etwa für Anwendungen der Informationsextraktion, des *question answering* oder der automatischen Textzusammenfassung. Unser Vorhaben geht von der These aus, dass diese oberflächennahe Verarbeitung mittlerweile an ihre Grenzen gestoßen ist. Wir streben daher an, Erkenntnisse der linguistischen Analyse wieder verstärkt für computerlinguistische Anwendungen nutzbar zu machen. Insbesondere soll eine Verbindung zwischen empirischen, korpusbasierten Arbeiten und ihrer computerlinguistischen Implementierung geschaffen werden.

Die linguistische Fragestellung, anhand derer dieses methodische Ziel verfolgt werden soll, betrifft sprachliche Mittel der kausalen Kohärenzstiftung (Kausalitätsmarker): Untersucht werden Konnektoren wie *weil, da, denn, sodass* und *deshalb*, kausal interpretierbare Präpositionen wie *wegen, infolge, durch* und *aus*, Substantive wie *Ursache, Folge, Wirkung* sowie verbale Einheiten wie *verursachen* oder *zur Folge haben*. Die Untersuchung stützt sich auf ein Korpus von Kommentarartexten. Die Marker sollen inventarisiert, hinsichtlich Syntax, Semantik, Informationsstruktur analysiert und ihr Beitrag zur Textkohärenz formalisiert werden. Die Ergebnisse sollen in die computerlinguistische Nutzung einmünden, die damit in die Lage versetzt wird, zusätzlich zu statistischen Informationen wieder verstärkt qualitative Informationen auszuwerten. Es wird ein prototypischer Diskursparser entwickelt, der Informationen aus Kausalitätsmarkern verarbeitet und unterspezifizierte Textstrukturen generiert. Die Erträge des Projektes können in einem späteren Schritt (außerhalb des hier beantragten Vorhabens) in Computeranwendungen für Aufgaben der automatischen Textzusammenfassung genutzt werden.