

Linguistische Einheiten im Schreibprozess. Systematische Untersuchung von Planungs- und Redigiereinheiten

Dr. phil. Cerstin Mahlow

Das Verfassen von Texten ist ein bedeutender Teil unserer alltäglichen und professionellen Kommunikation. Schreiben ist eine komplexe Aufgabe: zahlreiche kognitive, linguistische, technische und handwerkliche Anforderungen müssen gemeistert werden. Während linguistische Aspekte von «Text» ein etabliertes Untersuchungsgebiet innerhalb der Linguistik wie der Computerlinguistik sind, wird der *Prozess* der Entstehung linguistischer Einheiten wie Mehrwortausdrücke, Koreferenzketten und Diskursstrukturen bislang nicht untersucht.

In diesem Forschungsvorhaben untersuche ich systematisch die Produktion von diskursiven Mehrwortausdrücken anhand von Keystroke-Logging-Daten für ca. 90 Schreibsitzen. Pilotstudien zeigen, dass solche Einheiten auf eine Art und Weise produziert werden, die vermuten lassen, dass sie von Autoren als feste Einheit existieren. Die Resultate dieser Untersuchungen werden zudem dazu verwendet, detaillierte kontrollierte Experimente zur Untersuchung weiterer spezifischer linguistischer Einheiten zu entwerfen.

Keystroke-Logging-Daten von Schreibsitzen ermöglichen es, den Schreibprozess zu rekonstruieren und Prozess und Produkt aufeinander zu beziehen. Es kann nachvollzogen werden, wie bestimmte Phrasen entstanden sind, welche Änderungen vorgenommen wurden, ob Autoren tatsächlich das Hauptargument eines Textes als erstes notieren oder die Argumentationskette mehrfach reorganisiert wird. Keystroke-Logging-Daten werden sehr schnell sehr gross, so dass sie nur mit Hilfe automatischer Untersuchungsmethoden ausgewertet werden können. In diesem Forschungsvorhaben untersuche ich daher auf einer zweiten Ebene, welcher Grammatikformalismus aus dem Bereich der Computerlinguistik am geeignetsten für die Abbildung der inkrementellen Textproduktion ist. Die erwarteten Erkenntnisse des Forschungsprojektes erlauben für Text erstmals eine systematische und reproduzierbare Verbindung von Prozess und Produkt auf linguistischer Ebene.