

# 1 Abstract

Die Erstellung einer objektiven, reliablen Annotation für sprachlich komplexe Phänomene in rohsprachlichen Korpora ist eine schwierige Aufgabe, die bislang nur unzureichend gelöst worden ist.

Um die Objektivität einer Annotation zu gewährleisten, wird diese von mehreren Annotatoren durchgeführt. Unter Umständen muss — insbesondere bei geringer Übereinstimmung der Annotatoren — jedoch mitunter eine geringe Reliabilität der auf diese Weise ermittelten Annotation in Kauf genommen werden.

In der vorliegenden Arbeit wurde explorativ ein statistisches Verfahren zur Verbesserung der Reliabilität linguistischer Annotationen konzipiert. Die Interrater-Reliabilität (Beurteilerübereinstimmung) gibt zunächst einen Hinweis über die Güte der Übereinstimmung. Liegt eine geringe bis mittelmäßige Übereinstimmung vor, kann mittels Trennschärfekoeffizienten und Cronbach's Alpha die Korrelation der einzelnen Annotatoren mit dem Gesamturteil bzw. die Interne Konsistenz, d.h. die Homogenität der einzelnen Annotatoren/Beurteilungen sichtbar gemacht werden. Zusätzlich kann eine Faktorenanalyse durchgeführt werden, um die beobachteten Zusammenhänge zwischen den Beurteilern zu erklären. Annotatoren, die nicht homogen beurteilen, können aus der Beurteilung ausgeschlossen werden. Die Übereinstimmung - und damit die Güte der Annotation - wird optimiert.

Das Verfahren wurde in der vorliegenden Arbeit zur Identifikation und Annotation von Selbstgesprächanteilen im Dialog explorativ entwickelt.