

Gebärdensprachkorpora sind in der Korpuslinguistik insofern Exoten, als sie Sprachdaten von Sprachen enthalten, für die es keine Schriftform und damit auch keine Orthographie gibt. Der Schwerpunkt der Arbeit in aktuellen Korpusprojekten für Gebärdensprachen liegt in Verarbeitungs-schritten wie Segmentierung und Lemmatisierung, die üblicherweise weitgehend auto-ma-tisch ablaufen, hier jedoch aufwändig von Hand erledigt werden müssen. Da es für kaum eine Gebärdensprache eine umfangreiche lexikalische Datenbank gibt, sind viele in diesen Arbeits-schritten zu treffende Entscheidungen vorläufig. Dies macht es umso notwendiger, sich in den folgenden Arbeitsschritten nicht ausschließlich auf die gerade generierte Textform zu stützen, sondern immer Zugriff auf die Originaldaten zu haben. Wir präsentieren unseren Ansatz, der dieser Notwendigkeit Rechnung trägt und gleichzeitig verschiedene Sichtweisen auf die Daten bietet, um auch dann ein Maß an Qualitätssicherung zu erreichen, wenn die unabhängige Doppeltranskription aus Kostengründen ausscheidet.

Zur Illustration verwende ich Daten aus dem Map-Task des Projektes Dicta-Sign, in dem Datenerhebungen mit demselben Material für vier Gebärdensprachen durch-geführt worden sind. Dieses Beispiel zeigt auch, inwieweit Parallelen zwischen gebärden-sprachlichen und lautsprachlichen Korpusansätzen zu erwarten sind.