

Beispiele für Daten und Werkzeuge der CLARIN-Infrastruktur für die Geistes- und Sozialwissenschaften

CLARIN-D ist ein Forschungsinfrastrukturprojekt, das Ressourcen und Dienste anbietet, die WissenschaftlerInnen aus geistes- und sozialwissenschaftlichen Disziplinen in ihrer Forschung unterstützen (Siehe etwa die ESFRI-Broschüre der Europäischen Kommission, S. 4¹). Auf diesem Poster stellen wir einige Diensten und Ressourcen vor, die im Rahmen des CLARIN-Konsortiums für sprachwissenschaftliche Nutzer entwickelt wurden.

CLARIN-Service: Phonetische Alignierung, Konvertierung und Visualisierung mit WebMaus

WebMaus ist ein Werkzeug, mit dessen Hilfe man eine Segmentierung von Sprachsignalen automatisch vornehmen kann, wobei man dazu nur das Sprachsignal – eine Aufnahme – und eine Transkription benötigt. Forschende, die sich mit gesprochener Sprache beschäftigen, können so von den Transkriptionen zu den relevanten Stellen in der Aufnahme gelangen und weitere Analysen mit dem Werkzeug ihrer Wahl vornehmen. Wortgrenzen und Pausen werden automatisch erkannt. Als Teil der CLARIN-Infrastruktur ist es möglich, mit einem einheitlichen Login online auf dieses Werkzeug zuzugreifen und die Ergebnisse mit anderen CLARIN-Werkzeugen zu verbinden.

CLARIN-Ressourcen

Wir stellen Ressourcen vor, die innerhalb von CLARIN verfügbar sind, etwa das Saarbrücker Kochbuch-Corpus und historische, digital verfügbare Volltexte des Deutschen Textarchivs. Zudem werden Verfahren präsentiert, wie man Texte in die CLARIN-Infrastruktur aufnehmen und bearbeiten kann und wie man linguistisch annotierte Texte mit Hilfe der CLARIN-Infrastruktur weiter analysieren kann. Die Integration in die CLARIN-Infrastruktur erlaubt es, Korpora gesprochener und geschriebener Sprache sowie multimodale Daten mit einer Vielfalt von computerlinguistischen Werkzeugen zu analysieren. Dazu gehören u.a. Wortarten-Tagger, Lemmatisierer und die automatische Syntaxanalyse.

CLARIN-Suche nach linguistischen Ressourcen

Innerhalb von CLARIN gib es mehrere Möglichkeiten, nach Ressourcen zu suchen. Neben einer facettierten Suche im Virtual Language Observatory wird auch eine Suche mittels einer geografischen Karte angeboten. Linguistische Ressourcen werden mit den Orten, an denen sie entstanden sind oder aufbewahrt werden, auf einer Karte markiert, sodass man leicht nach Ressourcen mit Verbindung zu bestimmten Regionen suchen kann. Der Verweis auf den entsprechenden Eintrag in einem Ressourcenkatalog erlaubt einen ersten Eindruck, ob die Ressource näher betrachtet werden sollte. Auch auf diese Anwendung kann man über die CLARIN-Infrastruktur zugreifen. CLARIN ermöglicht zusätzlich einen direkten, webbasierten Zugriff auf zahlreiche Ressourcen für eine Vielzahl von Sprachen.

¹ http://ec.europa.eu/research/infrastructures/pdf/esfri_brochure_0113.pdf