

Eurac Research ist europäisches Zentrum für digitale Sprachdaten
Linguisten von Eurac Research beteiligen sich am Aufbau einer europäischen Datenbank zur
Nutzung von digitalen Sprachdaten für Forschung, Bildung und Kultur

Wie haben Medien in verschiedenen Ländern in den vergangenen Jahren über den Klimawandel berichtet? Verändert sich die deutsche Sprache in Südtirol anders als etwa in Österreich oder Deutschland? Hat das Schreiben in sozialen Netzwerken einen Einfluss auf das Schreiben in anderen Genres? Damit Wissenschaftler solchen – und unzähligen anderen – interessanten Fragen nachgehen können, benötigen sie große Mengen an Sprachdaten: Tonaufnahmen, Zeitungsartikel oder andere Texte. Damit einmal gesammelte Sprachdaten nicht verloren gehen und weiterhin genutzt werden können, hat die EU die Forschungsinfrastruktur CLARIN ins Leben gerufen: Sie ermöglicht es, die Daten in einem digitalen Archiv langfristig aufzubewahren und allen Interessierten zugänglich zu machen. Seit Kurzem ist Eurac Research mit seinem digitalen Sprachdatenschatz neben dem nationalen Koordinator in Pisa das zweite CLARIN-Zentrum Italiens.

„Der Aufbau einer solchen digitalen Forschungsinfrastruktur ist extrem wichtig. Wenn man mit dem Zug verreisen möchte, braucht es auch eine Infrastruktur, sprich es braucht Schienen, Bahnhöfe, die verschiedenen Länder müssen die gleiche Stromspannung haben, die gleiche Spurweite, die Zeiten müssen abgestimmt sein. Genau das Gleiche gilt, wenn wir Linguisten zum Beispiel Sprachdaten aus Deutschland analysieren wollen. In dem Fall müssen wir uns auch absprechen und die gleichen Formate verwenden, damit die Daten vergleichbar sind“, erklärt Andrea Abel, Leiterin des Instituts für Angewandte Sprachforschung von Eurac Research. „Die Infrastruktur ist insgesamt von großem Nutzen dabei, unser Engagement in den Digital Humanities zu verstärken. Die Digital Humanities wenden computergestützte Methoden auf die Geisteswissenschaften an und schaffen viele Schnittstellen für die Zusammenarbeit zwischen Forschungseinrichtungen und anderen Institutionen wie Bibliotheken, Archiven und Museen.“ Das Bozner Forschungszentrum zählt seit Kurzem zu den über ganz Europa verteilten CLARIN-Zentren. Die Linguisten von Eurac Research speisen nun ihre vorhandenen digitalen Sprachdaten nach CLARIN-Standards in die Webseite <http://clarin.eurac.edu> ein. Auch gibt es bereits Zusammenarbeiten mit anderen Institutionen, wie etwa der Landesbibliothek Friedrich Tessimann oder der Sprachstelle am Südtiroler Kulturinstitut, in denen erprobt wird, wie Sprachdaten gesammelt und genutzt werden können. Zudem können Daten von anderen interessierten Institutionen und Forschern in das Archiv aufgenommen werden. „Auf diese Weise entstehen große Mengen relevanter Sprachressourcen, die Antworten auf ganz neue Fragestellungen aus mehreren Disziplinen erlauben. Das Clarin-Netzwerk bietet eine zentrale Suchseite und die Wissenschaftler können die Daten leichter untereinander austauschen, neu kombinieren und vergleichen“, erklärt der Computerlinguist Alexander König, der CLARIN-Experte von Eurac Research.

Bozen, 29.05.2018

Kontakt: Stefanie Gius, stefanie.gius@eurac.edu, Tel. 0471 055 034, mobil 366 6633860