

DEUTSCH IN EUROPA

Sprachpolitisch, grammatisch, methodisch

METHODENMESSE

Mittwoch, 11. März 2020, 16:00 Uhr

KORAP UND EURECO - RECHERCHIEREN IN MEHRSPRACHIGEN VERGLEICHBAREN KORPORA

Nils Diewald, Franck Bodmer, Peter Harders, Marc Kupietz, Eliza Margaretha und Helge Stallkamp (IDS) sowie Elena Irimia (Romanian Academy Research Institute for Artificial Intelligence „Mihai Drăgănescu“)

Obwohl KorAP primär als Such- und Analyseplattform für DeReKo entwickelt wird, ist es sprachunabhängig konzipiert. Dies bedeutet zum einen, dass Korpora unterschiedlicher Sprachen mit beliebigen Annotationen durch KorAP zugänglich gemacht werden können und das Interface durch Lokalisierung/Internationalisierung sowohl für diese Korpora als auch für fremdsprachige Nutzer angepasst werden kann. Zum anderen ist KorAP unabhängig in Bezug auf Anfragesprachen und ermöglicht durch die Unterstützung von, unter anderem, COSMAS II, Poliqarp oder Annis, verschiedenen Nutzerkreisen einen leichten Einstieg in die Verwendung.

Neben einem großen Teil von DeReKo werden im Rahmen der EuReCo-Initiative bereits weitere fremdsprachige National- und Referenzkorpora durch KorAP zur Verfügung gestellt, darunter das rumänische Referenzkorpus CoRoLa (Barbu Mititelu et al., 2018) und Teile des ungarischen Nationalkorpus HNC (Oravecz et al., 2014). Um diese Korpora untereinander vergleichen zu können, bietet KorAP die Möglichkeit, Sub-Korpora nach vergleichbaren Kriterien in Form virtueller Korpora zu erstellen und bei der Suche und Analyse zu referenzieren.

Im Rahmen der Methodenmesse stellen wir die neuesten Entwicklungen von KorAP vor, unter anderem in Bezug auf statistische Erhebungen in vergleichbaren Korpora durch Verwendung eines neu entwickelten Web API Clients in R.

KorAP für DeReKo: <https://korap.ids-mannheim.de/>

KorAP für CoRoLa: <http://corola.racai.ro/>

Referenzen:

Barbu Mititelu, V./Tufiş, D./Irimia, E. (2018): The reference corpus of the contemporary romanian language (CoRoLa). In: Calzolari, N./Choukri, K./Cieri, C./Declerck, T./Goggi, S./Hasida, K./Isahara, H./Maegaard, B./Mariani, J./Mazo, H./Moreno, A./Odijk, J./Piperidis, S./Tokunaga, T. (Hg.), Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC'18), S. 1235–1239, Miyazaki/Paris. European Language Resources Association (ELRA).

Oravecz, C./Váradi, T./Sass, B. (2014): The Hungarian gigaword corpus. In: Calzolari, N./Choukri, K./Declerck, T./Loftsson, H./Maegaard, B./Mariani, J./Moreno, A./Odijk, J./Piperidis, S. (Hg.), Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), S. 1719–1723, Reykjavik/Paris. European Language Resources Association (ELRA).