Projects involved

- Archive of General Reference Corpora of Contemporary Written German
- Methods of Corpus Analysis and Corpus Mining

Researchers involved

- Dipl.-Ing. Cyril Belica
- Dr. Marc Kupietz
- Dr. Harald Lüngen
- Dipl.-Inf. Rainer Perkuhn

Contact

Dr. Marc Kupietz
Corpus Linguistics Programme Area
IDS – Institute for the German Language
Postfach 10 16 21
D-68016 Mannheim
Germany

Phone: 0621/1581-0
Fax: 0621/1581-200
Email: corpuslinguistics@ids-mannheim.de

INSTITUT FÜR
DEUTSCHE SPRACHE

R 5, 6–13
D–68161 Mannheim
www.ids-mannheim.de

Mitglied der
Leibniz-Gemeinschaft

The Institute for the German Language (IDS) is the central institution for the study and documentation of the contemporary usage and recent history of the German language. Together with 85 other non-university research institutes and service facilities, it belongs to the Leibniz Association, one of the four major research organisations.

INSTITUT FÜR
DEUTSCHE SPRACHE

Corpus Linguistics Programme Area

DeReKo

The German Reference Corpus

One of the main objectives of the IDS is to maintain an empirical basis for linguistic research with respect to contemporary written German. To this end, the institute has established in 1964 what has over time grown to become a large-scale electronic sample of contemporary German texts: the Mannheim German Reference Corpus (DeReKo). It is a declared IDS policy to provide for the long-term sustainability of this corpus archive.

DeReKo

- currently comprises 5.4 billion words and constitutes the largest linguistically motivated collection of contemporary German texts
- contains fictional, scientific and newspaper texts, as well as several other text types
- contains only licenced texts
- is encoded with rich meta-textual information
- is fully annotated morphosyntactically (three concurrent annotations)
- is continually expanded, with a focus on size and stratification of data
- may be analyzed free of charge via the query system COSMAS II
- serves as a "primordial sample" from which users may draw specialized sub-samples (so-called "virtual corpora") to represent the language domain they wish to investigate

Herzlichen Dank all denen, die uns die Texte für unser Korpus zur Verfügung stellen

Ludwig M. Eichinger