The project “Lexik des gesprochenen Deutsch” (LeGeDe) is a third-party funded project of the Leibniz Association (funding line: innovative projects) and was implemented from 2016 to 2019 as a cooperative project of the departments Pragmatics and Lexical Studies at the IDS. In the last years, both departments have gained the necessary expertise for the development of a corpus-based dictionary of spoken German, the conception of which is a first for German as well as other languages.

The following preconditions rendered this project possible:

(i) The FOLK-corpus (= “Research and Teaching Corpus of Spoken German”), developed at the IDS, constituted an appropriate data basis that could be used for the analyses.

(ii) Through project cooperations and active exchange, the LeGeDe-project could benefit from the expertise of the Department of Pragmatics in the area of semantic-functional analysis of spoken language in interaction.

(iii) In recent years, the areas “Computer Lexicography” and “Empirical Methods” could gain expertise in the conception and realization of complex multimedia online dictionaries as well as the development of relevant empirical methods. The LeGeDe-prototype ties in with this experience.

The following preconditions rendered this project possible:

(i) The FOLK-corpus (= “Research and Teaching Corpus of Spoken German”), developed at the IDS, constituted an appropriate data basis that could be used for the analyses.

(ii) Through project cooperations and active exchange, the LeGeDe-project could benefit from the expertise of the Department of Pragmatics in the area of semantic-functional analysis of spoken language in interaction.

(iii) In recent years, the areas “Computer Lexicography” and “Empirical Methods” could gain expertise in the conception and realization of complex multimedia online dictionaries as well as the development of relevant empirical methods. The LeGeDe-prototype ties in with this experience.

The following preconditions rendered this project possible:

(i) The FOLK-corpus (= “Research and Teaching Corpus of Spoken German”), developed at the IDS, constituted an appropriate data basis that could be used for the analyses.

(ii) Through project cooperations and active exchange, the LeGeDe-project could benefit from the expertise of the Department of Pragmatics in the area of semantic-functional analysis of spoken language in interaction.

(iii) In recent years, the areas “Computer Lexicography” and “Empirical Methods” could gain expertise in the conception and realization of complex multimedia online dictionaries as well as the development of relevant empirical methods. The LeGeDe-prototype ties in with this experience.

The project description

With the development of the LeGeDe-prototype, an entirely new way of describing and presenting language was created, from the perspective of spoken language research as well as from the lexicographic view. The different word articles describe both the lexical characteristics and the interactional functions of each headword in authentic conversation (cf. figure below). Transcript excerpts with audio substantiate the analyses.

Data basis and selection of headwords

The data basis for the development of the LeGeDe-prototype is FOLK (250h audio recordings/2.4 mio. tokens as from September 2019), which is the first and largest corpus of spoken German and is constantly being extended. FOLK is accessible via the DGD (= “Database for Spoken German”) and contains recordings and transcripts gathered in German-speaking areas within different private, institutional, and public settings. It offers various options for the analysis of spoken data in terms of linguistic and interactional characteristics.

The selection of suitable headword candidates for the LeGeDe-prototype was based on a direct quantitative comparison between FOLK and parts of DEREKO (= “German Reference Corpus”), which represents the written usage of the German language. In addition to one-word lemmata, manual qualitative analyses could help identify multi-part headword candidates (e.g. keine Ahnung [‘no idea’], guck mal [‘look’]), which have an interactional function as a whole.

Project outcomes

With the development of the LeGeDe-prototype, an entirely new way of describing and presenting language was created, from the perspective of spoken language research as well as from the lexicographic view. The different word articles describe both the lexical characteristics and the interactional functions of each headword in authentic conversation (cf. figure below). Transcript excerpts with audio substantiate the analyses.

As a result, the prototype mainly targets usage situations in scientific settings (language documentation). Perspectives for the future user for a lexicographic resource on spoken German, the determination of characteristics in spoken usage with a focus on the lexicon on all levels (form, content/function, situation etc.), the development of new possibilities of lexicographic representation, which take into account the functions of lexical items in interactional contexts, their visualisation, the development of new ways of describing highly contextualized lexical data in multimedia formats through the integration of language documentation in the form of transcripts and audio files, the creation of the tool Lexical Explorer, with which the user can create individual word profiles that can be searched and analyzed in regard to word form variation, co-occurrences, and metadata based on corpus data from FOLK and GeWiss (= “Spoken Academic Language”).
Research topic

Modern dictionaries of the German language are often based on written German. In the project “Lexik des gesprochenen Deutch” (“Lexis of spoken German”), for the first time a prototype of a corpus-based dictionary of spoken German was developed. For this, lexical characteristics typical of spoken German were described, analyzed and processed lexicographically on the basis of the FOLK-corpus, which was developed at “The Leibniz Institute for the German Language” (IDS). The resource is embedded into OWIDplus and is available online at https://www.owid.de/legede/. It presents lexical and functional characteristics of spoken German as used in interaction in private and institutional settings.

Project supervisors

Prof. Dr. Meike Meliss (IDS/Universidad de Santiago de Compostela), Dr. Christine Möhrs (IDS)

Associated projects

- FOLK (= “Research and Teaching Corpus of Spoken German”)
- Research units “Computer lexicography” and “Empirical Methods”

Associated employees

Katja Arens, Prof. Dr. Arnulf Deppermann, Prof. Dr. Stefan Engelberg, Dolores Lemmenmeier (née Batinic), Dr. Antje Nolting (née Töpel), Dr. Thomas Schmidt, Sarah Torres Cajo, student assistants, interns.

Project Contact

Dr. Christine Möhrs, Dept. of Lexical Studies
The Leibniz Institute for the German Language
legede@ids-mannheim.de
http://www1.ids-mannheim.de/lexik/
lexik-des-gesprochenen-deutsch.html

The Leibniz Institute for the German Language (IDS) is the central institute for research and documentation of the German language in its usage and recent history. It is one of the over 90 extramural research and service institutions in the Leibniz-Gemeinschaft (Leibniz Association).