

Collection Policy und Aufnahmekriterien der Forschungsdatenbank Lernertexte FD-LEX

Die webbasierte Forschungsdatenbank Lernertexte – kurz FD-LEX – archiviert Daten aus der Schreibforschung und stellt sie der wissenschaftlichen Öffentlichkeit zur Nachnutzung bereit. Die Datenbank wurde im Jahr 2021 in den *Verbund Forschungsdaten Bildung*, einem Zusammenschluss aus Forschungsdatenzentren aus dem Bereich der Bildungsforschung, integriert. Dadurch ist es Forschenden möglich, Textkorpora aus Lernertexten im Portal forschungsdatne-bildung.de aufzufinden. Zudem können Forschende, die eigene Daten bei FD-LEX archivieren möchten, diese über das Portal melden. Im Zuge dieser Kooperation wurde die Datenbank ausgebaut und es werden nach und nach Korpora aus weiteren Schreibforschungsprojekten in FD-LEX aufgenommen.

Welche Daten werden in FD-LEX aufgenommen?

Im Folgenden werden Kriterien aufgeführt, die für eine Aufnahme erfüllt sein müssen.

Formale Kriterien

- Formale Kriterien betreffen die **Qualität der Dokumentation** als wichtigste Grundvoraussetzung für die Aufnahme der Studie.
- Es werden grundsätzlich nur Daten und Instrumente **abgeschlossener Studien** (bei längsschnittlichen Studien auch nach Abschluss von einzelnen Erhebungswellen) aufgenommen. Die Textdaten sollten **vollständig** vorliegen.
- Darüber hinaus ist die Übermittlung der **Begleit- und Kontextmaterialien** wie die verwendeten Schreibaufgaben sowie weitere relevante Erhebungsinstrumente für eine Aufnahme von Daten eine wesentliche Voraussetzung. Diese Materialien ermöglichen eine umfassende Kontextualisierung und stellen somit eine Grundvoraussetzung dar, um die Textdaten nachvollziehbar und damit nachnutzbar zu machen.
- Grundsätzlich gilt, dass die Daten einer Studie nur dann für Sekundäranalysen bereitgestellt werden können, wenn die **rechtlichen Rahmenbedingungen** dies erlauben. Die Nutzungsrechte müssen vorliegen und die Bedingungen der Datenweitergabe geklärt sein.

Inhaltliche bzw. dokumentarische Kriterien

Folgende (Meta-)Daten müssen für eine Aufnahme vorliegen:

- Eindeutige Identifikationsnummer für jede Person
- Alter (in Monaten oder Jahren) ODER Jahrgangsstufe
- Eindeutige Bezeichnung der Textsorte(n)

- Textlänge (Anzahl geschriebener Wörter pro Text)
- Texte transkribiert mitsamt Transkriptionsregeln
- Originaltexte, falls vorhanden (handschriftlich, als Scan)
- Geschlecht
- Sprache der Texte

Weitere wünschenswerte (Meta-)Daten sind:

- Sprachbiografie
- Sozioökonomischer Hintergrund
- Schulform
- Deutschnote
- Gruppenklassifikator (z. B. Nummer der einzelnen Schulklassen), falls zutreffend
- Gruppe im Forschungsdesign (z. B. Interventions- vs. Kontrollgruppe), falls zutreffend
- Messzeitpunkt, falls zutreffend
- Angabe digital oder analog, falls zutreffend

Für die Aufnahme werden zudem Texte zu Forschungszusammenhang und -design, Ablauf der Datenerhebung sowie Metadaten zu Forschungsprojekt und Studie benötigt.

Technische Kriterien

Folgende technische Kriterien müssen für eine Aufnahme erfüllt sein:

- Transkripte der Texte: pro Text eine Datei
- Scans der Originaltexte: pro Text eine Datei, Benennung nach Identifikationsnummer, im PDF-Format
- Metadaten: in einer Excel-Datei und eindeutige Zuordnung zu Identifikationsnummer

Der Import der Daten ist auf zwei unterschiedliche Arten möglich, die sich nach der Größe und Komplexität des bereitgestellten Textkorpus richten. Je nachdem müssen die Daten für die Aufnahme auf die eine oder andere Art aufbereitet werden. Im Falle der Passung neuer Daten wird dies über individuelle Absprachen festgelegt.

Allgemein gilt für den Import der Daten, dass keine Metadaten aus Textdateien ausgelesen werden.

Die konkreten Anforderungen auf inhaltlicher und technischer Ebene und die mögliche Darstellung in der Datenbank werden spezifisch für das betreffende Textkorpus abgestimmt.

Wenden Sie sich für Fragen zur Eignung Ihres Textkorpus und individuelle Abstimmungen per E-Mail an Sabine Hachmeister und Sandra Tietjens unter der Adresse: [FD-LEX-MI\(at\)uni-koeln\(dot\)de](mailto:FD-LEX-MI(at)uni-koeln(dot)de).

Eine erste Orientierung bietet diese [Übersichtsseite zur Datenbank FD-LEX](#). Auf der [Datenbankseite](#) selbst kann nach der Registrierung zudem eingesehen werden, wie die Daten den Nutzenden bereitgestellt werden.