TEXTDATEN ROMANISTIK

Textdaten Romanistik

Ein Wiki zur Arbeit mit digitalen Ressourcen in den romanischen Sprachwissenschaften

Blended Learning-Konzept

- Präsenzveranstaltungen
  - Innerhalb und außerhalb von Seminaren
  - Zielgruppenspezifisch: Teilbereich oder allgemeine Einführung
  - Fortsetzung im Selbststudium
- Sprechstundenunterstützung
  - Unterstützung der Sprechstunde: Hausarbeiten, Abschlussarbeiten
  - Fortsetzung im Selbststudium
  - Rückfragen in der Sprechstunde
- Nutzer von außen
  - Selbststudium

TEXT DATEN

Seiten / Textdaten Romanistik Home

Bearbeiten Beobachten Teilen Extras

Selbstlernmaterialien zum Wiki

Der bewusst weit gefasste Begriff *Textdaten* verweist auf die diversen Arten von *Texten* - Korpora, historische und aktuelle Literatur, Wörterbücher, Sprachatlanten, Sekundärliteratur - die als Informationsquelle oder als direkter Untersuchungsgegenstand zur Klärung einer Fragestellung beitragen können.

Von *Daten* ist die Rede, da es sich um digitalisierte Texte handelt, die dem Benutzer online zugänglich sind.

Textdaten können mithilfe entsprechender Software digital analysiert und weiterverarbeitet werden.

werden. Eine genauere Erläuterung zur Beschaffenheit von Korpora und der zielgerichteten Arbeit mit ihnen ist in der Subkategorie Grundlagen der Korpusarbeit zu

Damit bilden sie eine Datengrundlage, auf der sich sprachliche Merkmale differenziert und statistisch aussagekräftig untersuchen lassen.

Aufschluss über weiter zurückliegende Sprachstadien geben und damit auch zum Objekt linguistischer Untersuchungen werden können. Sie stehen in Online-Bibliotheken und Datenbanken zur Verfügung.

In den Bereichen Wörterbücher und Sprachatlanten kann der Nutzer sich über lexikalische Besonderheiten einzelner Sprachen, Sprachstufen und Varietäten informieren und sich ein Bild von der regionalen Ausprägung der Sprachen der Romania machen.

Zur Unterstützung der selbständigen Informationssammlung im Internet und der Erschließung unterstützender Forschungsliteratur verweist das Wiki Textdaten Romanistik im entsprechenden Bereich auf einschlägige bibliographische, biographische und enzyklopädische Rechercheportale.

Wer sich an die aktive Arbeit mit den Textdaten wagen will, und vielleicht sogar selbst Korpora erstellen oder Texte zur Analyse aufbereiten möchte, findet im Bereich Tools diverse gängige Programme zur Transkription und Annotation.

Welche Ziele werden verfolgt?

- Bereitstellen von Sprachdaten und Analysetools
- Heranführen an die Arbeit von Textdaten mithilfe praxisorientierter Manuals
- Liste von linguistisch relevanten Konventionen
- Förderung der computergestützten Arbeit mit Daten
- Entlastung von Seminaren und Sprechstunden: Antworten zu häufigen Fragen
- Einbindung der Studierenden: Evaluation und Übungen

Inhalte des Wikis

- Das Wiki selbst enthält keine Textdaten, Ressourcen werden
  - verlinkt
  - geordnet
  - beschrieben
- Das Wiki enthält weiterhin
  - Links zu Programmen, die der Arbeit mit den Textdaten dienen
  - Anleitungen für diese Programme
  - Selbstlernmaterialien
  - Übungen

Wie findet man im Wiki, was man sucht?

Ein detailliertes **Schlagwortsystem** ermöglicht die thematische Suche nach benötigten Ressourcen:

- Art der Ressource *(corpus, database, dictionary, tool…)*
- Sprache *(portuguese, catalan, latin, anglo-norman…)*
- Textsorte *(interviews, journalism, literature, theatre, language_and_law…)*
- Historische Einordnung *(1300, old_french, renaissance, middle_ages, diachrony…)*
- Sprachliche Besonderheiten *(dialects, child_language, written_language…)*
- Aufbereitung/Analysemöglichkeiten *(part_of_speech, lemmatized…)*

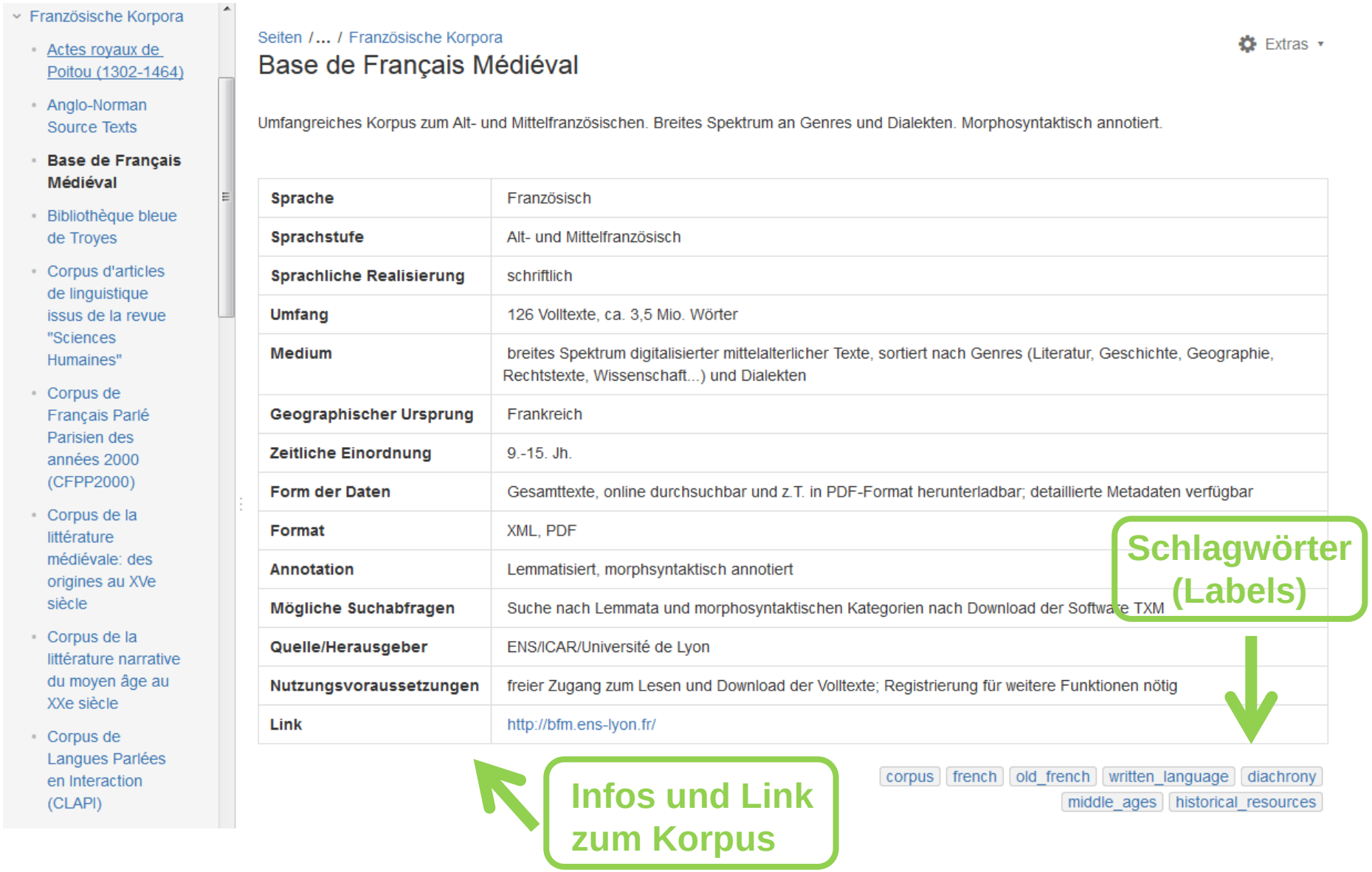
Häufig gesuchte Labels

Labels

Bereiche: 1. Linguistische Korpusarbeit

- Mithilfe von Korpora kann man sehr **große Datenmengen** zuverlässig analysieren
- Sprachliche Merkmale werden computerlesbar und **zählbar** gemacht
- **Tools und Anwendungen** helfen bei der Annotation und Auswertung von Texten.
- Auf Basis der **Häufigkeit charakteristischer Merkmale** ist eine Analyse von grammatischen Strukturen, die Beschreibung von Sprachregistern, Varietäten, Sprachwandel usw. möglich

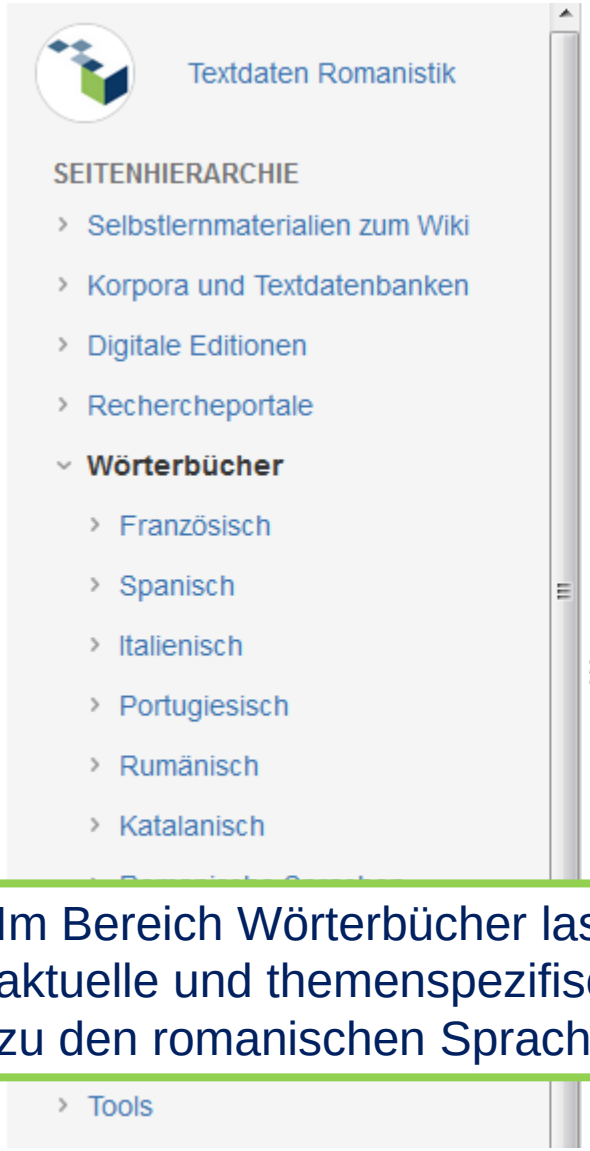
Bereiche: 1. Linguistische Korpora

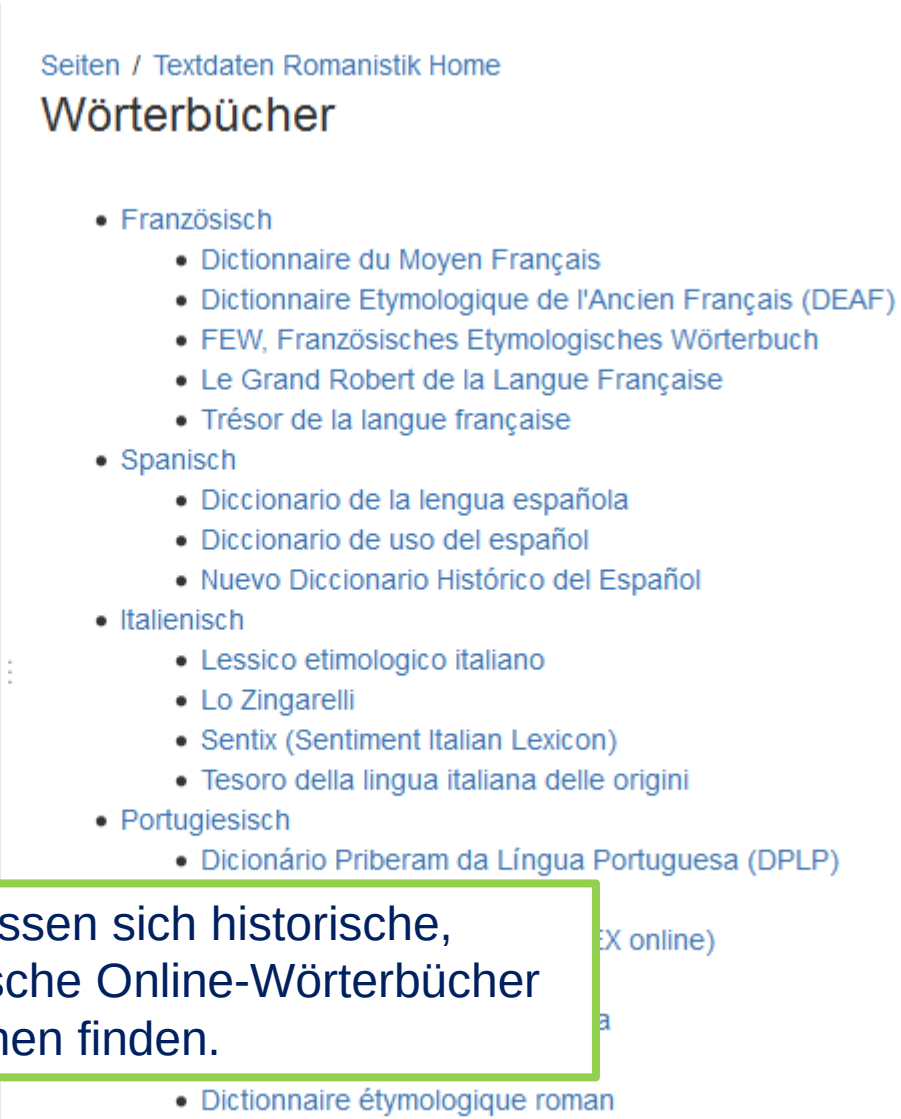


Bereiche: 2. Rechercheportale

Unter den Begriff Rechercheportale fallen Online-Bibliographien, Suchmaschinen und Enzyklopädien, die bei der Suche nach Fachliteratur und Fachwissen helfen.

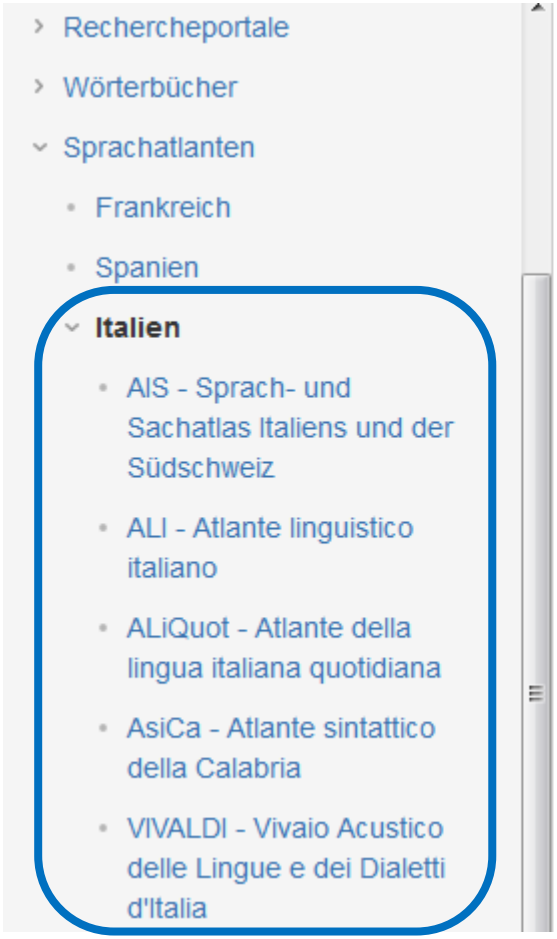
Bereiche: 3. Wörterbücher





Im Bereich Wörterbücher lassen sich historische, aktuelle und themenspezifische Online-Wörterbücher zu den romanischen Sprachen finden.

Bereiche: 4. Sprachatlanten



Online-Sprachatlanten informieren mithilfe von Karten und Audiomaterial über sprachliche und dialektale Besonderheiten bestimmter Regionen.

Bereiche: 5. Tools

Im Bereich Tools befindet sich eine Liste von Programmen, die für die Arbeit mit Textdaten gebraucht werden. Zu jedem Programm gibt es eine Kurzbeschreibung und den Link zur Installation.

Bereiche: 6. Manuals

SEITENHIERARCHIE

- Korpora und Textdatenbanken
- Digitale Editionen
- Rechercheportale
- Wörterbücher
- Sprachatlanten
- Manuals
  - Audiotranskription
    - Praat Kurzanleitung
    - Tastaturfreundliche phonetische Zeichen: SAMPA statt IPA
    - **Transcriber Kurzanleitung**
  - Datenkodierung und XML
  - Digitale Textedition mit TEI
  - Programmieren in der Linguistik
  - Selbstlernmaterialien zum Wiki
  - Wissenschaftliches Arbeiten, allgemein
- Tools

Arbeiten mit Transcriber

Das Programm ist im Internet frei verfügbar (http://sourceforge.net/projects/trans/files/transcriber/1.5.1/).
Installation entsprechend den Anweisungen in der *Readme*-Datei.

Das Programm erzeugt ein Ikon auf dem Desktop. Wenn es angeklickt wird, startet das Programm.

Zunächst kann eine eigene Tondatei (Format *.wav) geladen werden: "File", "New trans", im Explorerfenster Ablageort der entsprechenden *.wav-Da

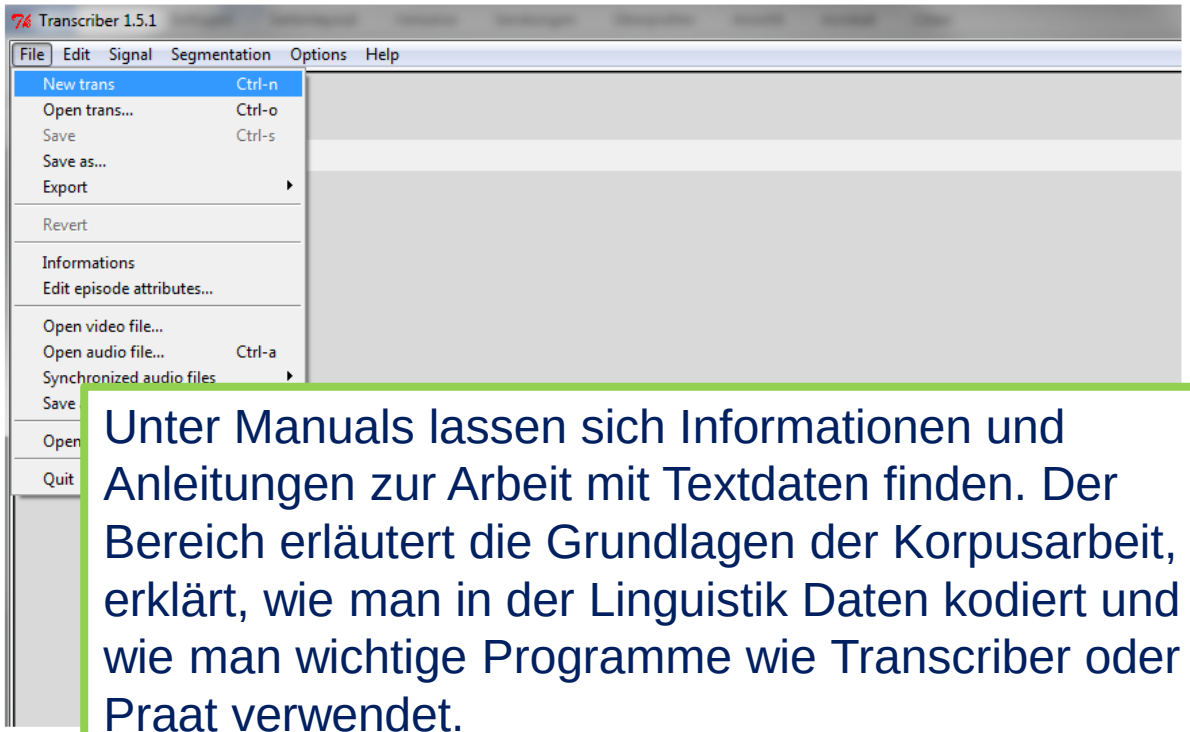
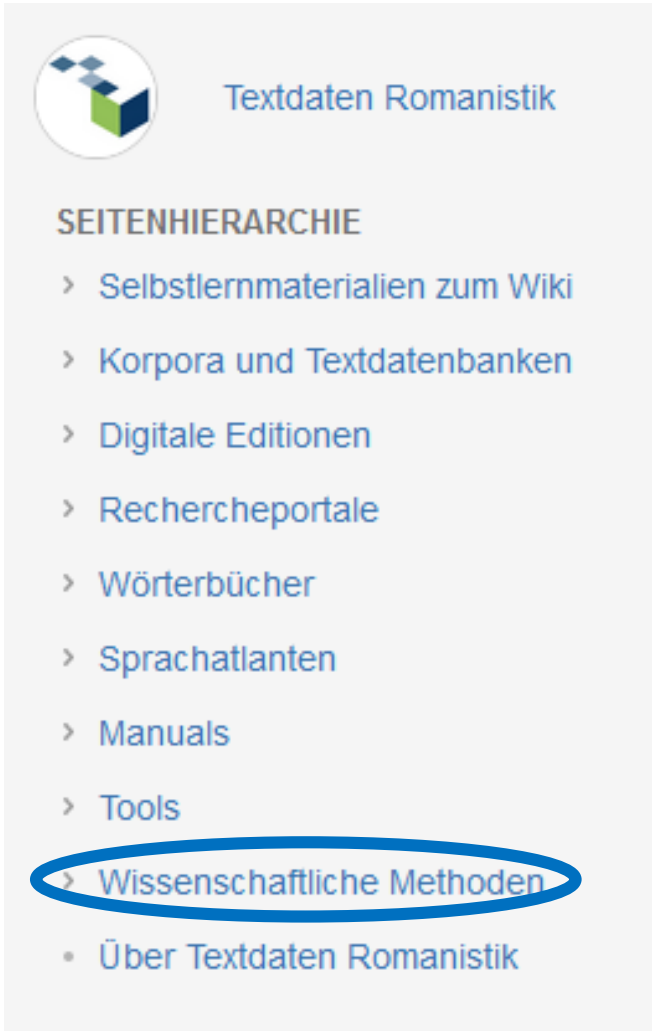


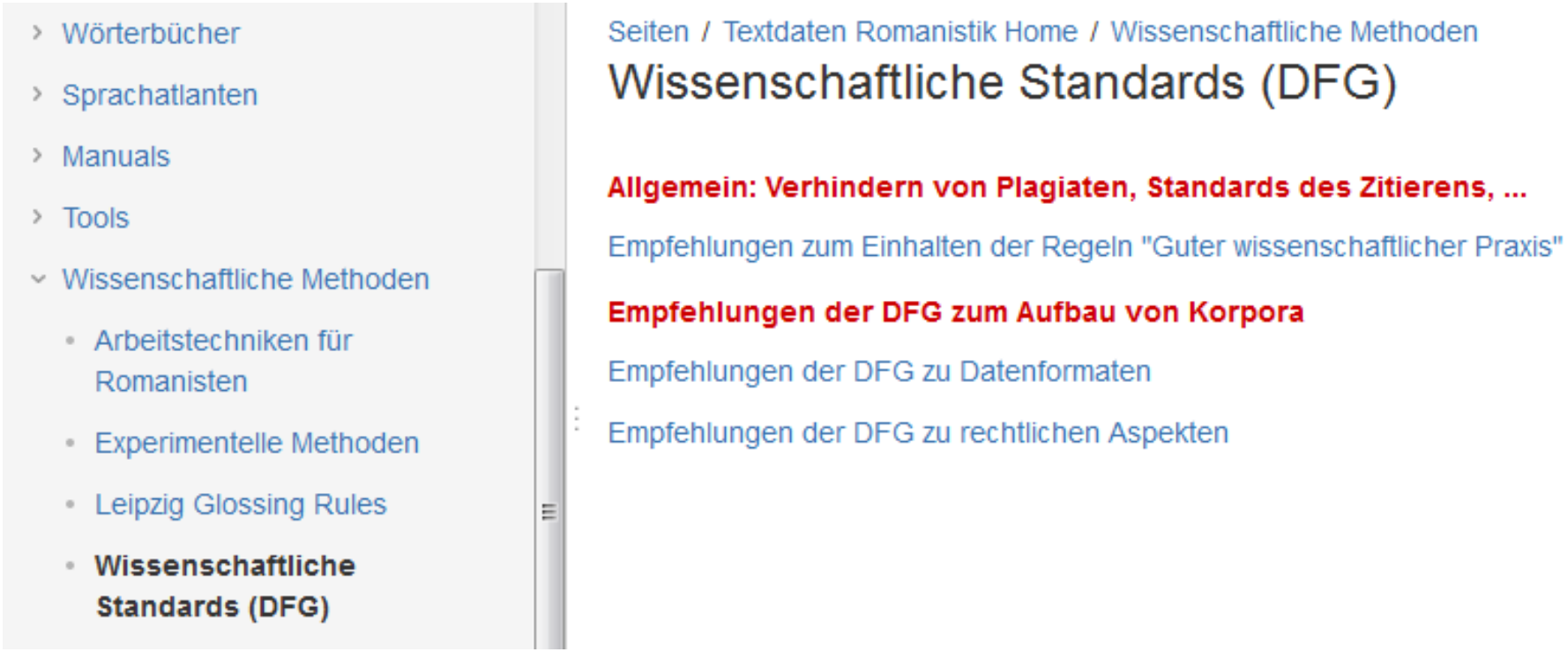
Fig. 1 Erstellen einer neuen Transkription

Unter Manuals lassen sich Informationen und Anleitungen zur Arbeit mit Textdaten finden. Der Bereich erläutert die Grundlagen der Korpusarbeit, erklärt, wie man in der Linguistik Daten kodiert und wie man wichtige Programme wie Transcriber oder Praat verwendet.

Bereiche: 7. Wissenschaftliche Methoden



Bereiche: 7. Wissenschaftliche Methoden



In diesem Bereich befinden sich Informationen zu den Grundlagen wissenschaftlichen Arbeitens: dazu, wie man erfolgreich eine wissenschaftliche Arbeit schreibt und welche Regeln beim Durchführen von Experimenten, beim Zitieren oder Glossieren gelten.

Relevante Literatur

Gerstenberg, Annette (^{2}2013): *Arbeitstechniken für Romanisten. Eine Anleitung für den Bereich Linguistik*. Berlin/Boston: De Gruyter (speziell Kap. 6).

HSK 29.1 = Lüdeling, Anke / Kytö, Merja (eds.) (2008): *Corpus Linguistics. An International Handbook*. Berlin/New York: De Gruyter.

Lemnitzer, Lothar / Zinsmeister, Heike (eds.) (2006): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.

Pusch, Claus D. / Raible, Wolfgang (2002): *Romanistische Korpuslinguistik: Korpora und gesprochene Sprache*. Tübingen: Narr (ScriptOralia, 126).

Barbera, Manuel / Corino, Elisa / Onesti, Cristina (eds.) (2007): *Corpora e linguistica in rete*. Perugia: Guerra.

Bilger, Mireille (1996): „Corpus de Portugais et d'Espagnol". In: *Recherches sur le français parlé* 1/2, 124-130.

Freddi, Maria (2014): *Linguistica dei corpora*. Roma: Carocci (Bussole, 490).

Williams, Geoffrey (ed.) (2005): *La linguistique de corpus*. Rennes: Presses universitaires de Rennes.