

## EMLex-Aufbaumodul A6

Modulbezeichnung	Aufbaumodul A6: Computerlexikographie	Gesamtzahl ECTS-Punkte
Lehrveranstaltungen	Blockseminar: 4 bis 5 Tage, 26-28 Unterrichtsstunden (entspricht 2 SWS)	5
Lehrende	Stephanie Evert (FAU), Ulrich Heid (HI), Besim Kabashi (FAU)	
Modulverantwortliche/-r	Stephanie Evert	
Inhalt	<p>Grundlagen der Korpuslinguistik</p> <ul style="list-style-type: none"> <li>• Ansätze und Methoden zur Auswertung von Korpora</li> <li>• Lexikographische Anwendungen von Korpusdaten</li> <li>• Korpus-Typologie, Übersicht wichtiger verfügbarer Korpora</li> <li>• Korpusdesign, Repräsentativität, Textquellen, Metadaten</li> </ul> <p>Erstellung eigener Korpora</p> <ul style="list-style-type: none"> <li>• Korpuserstellung aus Online-Quellen per Web-Scraping etc.</li> <li>• Bereinigung und Normalisierung, Gewinnung von Metadaten</li> <li>• Repräsentations- und Austauschformate (Standards)</li> <li>• Online- und Offline-Werkzeuge für Web-Scraping etc.</li> <li>• Automatische linguistische Annotation (Wortarten = POS, Lemmatisierung, Eigennamen, syntaktische Analyse, ...)</li> <li>• Online- und Offline-Werkzeuge für die automatische linguistische Annotation</li> </ul> <p>Suche in Korpora</p> <ul style="list-style-type: none"> <li>• Reguläre Ausdrücke</li> <li>• Zeichensätze, Unicode-Standard</li> <li>• CQP-Anfragesprache zur Suche nach lexikogramatischen Mustern</li> <li>• Übungen mit SketchEngine und CQPweb</li> </ul> <p>Quantitative Auswertungen</p> <ul style="list-style-type: none"> <li>• Häufigkeitslisten und Metadaten-Verteilung</li> <li>• Kollokationen und Word Sketches</li> <li>• Keyword-Analyse</li> <li>• Lexikographische Interpretation der Ergebnisse</li> <li>• Grundprinzipien der statistischen Inferenz</li> </ul> <p>Reproduzierbarkeit</p> <ul style="list-style-type: none"> <li>• Forschungsmethodik und Dokumentation</li> <li>• Datenmanagement, Nachhaltigkeit von Korpora und Ergebnissen</li> </ul>	
Lernziele und Kompetenzen	<p>Die Studierenden sollen in die Lage versetzt werden,</p> <ul style="list-style-type: none"> <li>• ihren Bedarf an Korpusdaten für ein lexikographisches Projekt zu formulieren und ein geeignetes Korpus zu konzipieren;</li> <li>• ein solches Korpus aus Webseiten oder anderen Quellen zu erstellen;</li> </ul>	

	<ul style="list-style-type: none"> <li>• das Korpus mit Hilfe automatischer Werkzeuge linguistisch zu annotieren;</li> <li>• im Korpus mit regulären Ausdrücken und komplexeren Suchanfragen nach lexikogrammatistischen Mustern zu suchen;</li> <li>• quantitative Methoden wie Kollokations- und Keyword-Analysen anzuwenden und die Ergebnisse in angemessener Weise zu interpretieren;</li> <li>• ihre Ergebnisse in Kurzvorträgen zu vermitteln;</li> <li>• wissenschaftliche Diskussionen über technische und methodologische Aspekte der Korpusarbeit zu führen;</li> <li>• ihre Korpora und Auswertungsergebnisse zu dokumentieren und zu archivieren.</li> </ul>
Voraussetzungen für die Teilnahme	25 ECTS-Punkte aus dem ersten Semester
Einpassung in Studienverlaufsplan	Wahlpflichtmodul im zweiten Semester (6 bzw. 7 von 10)
Verwendbarkeit des Moduls	für den Masterstudiengang EMLex
Studien- und Prüfungsleistungen	Die Dozenten wählen eine der folgenden Möglichkeiten: a) 90-minütige Klausur über den Inhalt des Seminars oder b) Kurzreferat im Seminar der Klasse plus kurze Hausarbeit (10-15 Seiten) oder c) längere Hausarbeit (15-20 Seiten)
Berechnung der Modulnote	100% Note der Modulabschlussprüfung a), b) oder c)
Turnus des Angebots	jährlich, nur im Sommersemester
Wiederholung der Prüfung	1
Arbeitsaufwand in Zeitstunden	z.B. Präsenzzeit: 28 Stunden (2 SWS mal 14 Wochen); Eigenstudium: 122 Stunden
Dauer des Moduls	Blockveranstaltung (Termin und Dauer nach besonderer Bekanntgabe)
Unterrichts- und Prüfungssprache	Deutsch und/oder Englisch
Literaturhinweise	HSK 5.4, Kap. XVIII + XIX
letzte Änderung	Juli 2024

Module name	Advanced module A6: Computational Lexicography	ECTS
Courses	Block seminar: 4 to 5 days, 26 to 28 teaching hours (equals 2 hours/week)	5
Lecturers	Stephanie Evert (FAU), Ulrich Heid (HI), Besim Kabashi (FAU)	
Module coordinator	Stephanie Evert	
Content	<p>Foundations of corpus linguistics</p> <ul style="list-style-type: none"> <li>• Principles and methods of corpus analysis</li> <li>• Applications of corpus data in lexicography</li> <li>• Types of corpora, overview of existing corpora</li> <li>• Corpus design, representativity, data sources, metadata</li> </ul> <p>Corpus compilation</p> <ul style="list-style-type: none"> <li>• Building corpora from online data: Web scraping etc.</li> <li>• Boilerplate removal, normalization, metadata extraction</li> <li>• Representation and exchange formats</li> <li>• Online and stand-alone tools for Web corpus compilation</li> <li>• Automatic linguistic annotation (POS, lemma, NER, parsing, ...)</li> <li>• Online and stand-alone tools for linguistic annotation</li> </ul> <p>Searching corpora</p> <ul style="list-style-type: none"> <li>• Regular expressions</li> <li>• Character encodings and the Unicode standard</li> <li>• CQP query language for lexico-grammatical patterns</li> <li>• Practical exercises with SketchEngine and CQPweb</li> </ul> <p>Quantitative analysis</p> <ul style="list-style-type: none"> <li>• Frequency lists and metadata distribution</li> <li>• Collocations and word sketches</li> <li>• Keyword analysis</li> <li>• Lexicographic interpretation of results</li> <li>• Foundations of statistical inference</li> </ul> <p>Reproducibility</p> <ul style="list-style-type: none"> <li>• Research methodology and documentation</li> <li>• Data management, sustainability of corpus resources</li> </ul>	
Aims and skills	<p>The students should be able</p> <ul style="list-style-type: none"> <li>• to formulate their corpus requirements for a lexicographic project and specify the design of a representative corpus;</li> <li>• to compile such a corpus from Web pages or other sources;</li> <li>• to annotate the corpus with linguistic information using automatic natural language processing tools;</li> <li>• to search the corpus with regular expressions and more complex queries based on lexico-grammatical patterns;</li> <li>• to apply quantitative techniques such as collocation or keyword analysis and interpret the results appropriately;</li> <li>• to communicate the results of their work to fellow students;</li> </ul>	

	<ul style="list-style-type: none"> <li>to lead academic discussions about technical and methodological aspects of corpus-based research; and</li> <li>to document and archive corpus data and analysis results.</li> </ul>
Prerequisites	25 ECTS from the first semester
Integration in curriculum	Elective module in the second semester (6 or 7 from 10).
Courses of study for which the module is acceptable	For the Master degree programme EMLex
Assessment and examinations	<p>The teachers choose one of the following:</p> <p>a) 90-minute final written exam on the contents of the seminar or</p> <p>b) presentation in class plus a brief paper (10-15 pages) or</p> <p>c) longer paper (15-20 pages)</p>
Grading procedure	100% of the mark obtained in a), b) or c).
Exam retake	1
Module frequency	Annually, only in summer term
Workload	e.g. class time: 28 hours (equals 2 hours multiplied by 14 weeks); study time: 122 hours.
Module duration	Block seminar (date and duration to be announced)
Course and examination language	German and/or English
Recommended reading	HSK 5.4, Ch. XVIII + XIX
Last changes	July 2024

