

DIGI / LING

TRANS / EUROPEAN / E-LEARNING

HUB / FOR / DIGITAL / LINGUISTICS

learn.digiling.eu STUDY WITH US

www.digiling.eu

VISIT US



Erasmus+ Key Action 2 (KA2): Strategic Partnership for Higher Education

DIGI

LING

ABOUT

DigiLing is an Erasmus+ Strategic Partnership project (2016-2019) which designed a model curriculum in digital linguistics to serve as a road map for the evolution of European university programmes which will conform to industry demands. To respond to these demands, the project carried out a labour market needs analysis among European companies. A set of key digital competencies was identified and a model curriculum for **Digital Linguistics** at Masters level was designed.

WHAT IS **DIGITAL LINGUISTICS**

Our vision of the profile of a Digital Linguist is interdisciplinary in that it entails skills in digital communication, multilinguality, language processing and information science. Our proposed curriculum is tailored to students with diverse backgrounds, from language and linguistics, humanities, social sciences, to information and data science or engineering. The differences between incoming students are alleviated through obligatory elective foundations in linguistics, information science and digital communication. The programme is modular with a central focus on digital skills and a high extent of elective courses covering various aspects of language technologies, intelligent applications, digital media, multilingual content creation and digital ethics.

SEMESTER

1st / 2nd

3rd

4th

FOUNDATIONS

METHODS & TOOLS

RESEARCH &
APPLICATIONS

LINGUISTICS	MULTILINGUAL COMMUNICATION
PROGRAMMING & IT	DIGITAL MEDIA

DIGITAL CONTENT AUTHORING	TEXT ANALYSIS
STATISTICS	ETHICS & LAW

APPLICATION-ORIENTED TECHNIQUES
PROJECT I / INTERNSHIP

SPECIAL TOPICS IN: Linguistics / Multilingual Communication / Programming & IT / Digital Media
--

ELECTIVE TOPICS FROM DIGITAL HUMANITIES
FREE ELECTIVES

PROJECT II

RECENT ADVANCES IN DIGITAL LINGUISTICS
--

RESEARCH & MA THESIS

DIGILING COURSES

Once the target skills of the future digital linguist had been established the DigiLing project proceeded with designing the e-learning hub with online courses covering some of the key areas in Digital Linguistics. Each consortium partner proposed to design one or two courses in accordance with their main scientific expertise and teaching experience.

The courses were designed in line with a common DigiLing framework consisting of the following principles:

- In terms of level, time commitment and study effort each course is equivalent to a face-to-face university course with 3-5 ECTS
- All courses employ various modes of content presentation and a range of interactive elements to facilitate learning
- Each course progressively guides the student towards achieving the learning outcomes, with regular feedback provided at different stages of the course
- All courses comply with open e-learning standards and are available free of charge upon registration through the DigiLing e-learning platform.

While the courses represent stand-alone learning tracks and may be taken individually, the conveyed skills and competencies are intricately interconnected and lead towards an advanced understanding of digital communication, multilinguality and language processing.

LIST OF DIGILING COURSES:

Variability of Languages in Time and Space

Univerzita Karlova, course instructors:
Zdeněk Žabokrtský, Anna Nedoluzhko,
Magda Ševčíková, Šárka Zikánová

Introduction to Text Processing and Analysis

Univerzita Karlova, course instructor: Pavel Vondříčka,
Lucie Chlumská

Post-Editing Machine Translation

Johannes Gutenberg Universität Mainz, course
instructors: Jean Nitzke, Anke Tardel

Introduction to Python for Linguists

Sveučilište u Zagrebu, course instructor: Petra Bago

Computational Lexicology and Lexicography

Sveučilište u Zagrebu, course instructor:
Nives Mikelić Preradović

Localization Tools and Workflows

University of Leeds, course instructor: Caroline Reiss

Mining and Managing Multilingual Terminology

Univerza v Ljubljani, course instructor: Špela Vintar

VARIABILITY OF LANGUAGES IN TIME AND SPACE

distinguish language types

identify writing system

understand diachronic language change

INTRODUCTION TO PYTHON FOR LINGUISTS

build complex CQL queries

use regular expressions

implement Python syntax and semantics

identify and implement variables, operators and functions and data structures in Python

identify and implement control flow in Python

MINING AND MANAGING MULTILINGUAL TERMINOLOGY

recognise terms and their properties in different settings

plan, design and populate a termbase in any chosen terminology tool

compile a specialised corpus of texts for terminology mining

use a corpus workbench to extract terms and definitions

understand what post-editing is and how to perform it

evaluate factors influencing post-editing

understand how machine translation works

POST-EDITING MACHINE TRANSLATION

apply knowledge of language diversity to language technology

distinguish data formats, their syntax and properties

understand and apply linguistic annotation on different levels

use corpus tools to analyse existing and own corpora

INTRODUCTION TO TEXT PROCESSING AND ANALYSIS

understand the content and limitations of print dictionaries for computational purposes

compare various kinds of mono- and multilingual subcategorisation lexicons

compare the design and content of various kinds of sentiment lexicons

COMPUTATIONAL LEXICOLOGY AND LEXICOGRAPHY

describe different tools to support the localization process

combine translation and other tools and resources to process language

use CAT tools

demonstrate awareness of the planning and management activities involved in the localization process

distinguish between human translation, post-editing and proofreading

LOCALIZATION TOOLS AND WORKFLOWS

COURSE TITLE

VARIABILITY OF LANGUAGES IN TIME AND SPACE

COURSE INSTRUCTORS

ANNA NEDOLUZHKO's main research interests are linguistic analysis of grammar, with special focus on phenomena exceeding the sentence boundary, semantic analysis of the verbal system and linguistic typology.

nedoluzko@ufal.mff.cuni.cz

MAGDA ŠEVČÍKOVÁ focuses on inflectional and derivational morphology; she took part in the Prague Dependency Treebank project and, recently, is involved in building the lexical network DeriNet for Czech.

sevcikova@ufal.mff.cuni.cz

ŠÁRKA ZIKÁNOVÁ deals with the analysis of syntax, discourse relations and topic-focus articulation in the Prague Dependency Treebank.

sarka.zikanova@mff.cuni.cz

Anna Nedoluzhko, Magda Ševčíková, and Šárka Zikánová are senior research associates at the Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic.

COURSE DESCRIPTION	The course provides students with basic information about the diversity of natural languages around the globe and the main dimensions along which they differ.
COURSE OBJECTIVES	<ul style="list-style-type: none"> • to understand basic linguistic notions needed for analyzing language variance on different levels of language description • to distinguish language types according to different typological criteria such as morphology and word order • to identify main types of writing systems • to understand the main principles of phonological, morphological and syntactic language changes from the diachronic point of view • to analyze implications of language diversity for contemporary language technology
TOPICS COVERED	<p>Languages around the world</p> <p>Linguistic sign, language system</p> <p>Grammar, lexicon, and word formation</p> <p>Word formation across languages</p> <p>Linguistic typology of grammar: Morphology</p> <p>Linguistic typology of grammar: Syntax</p> <p>Writing systems around the world</p> <p>Linguistic typology of grammar: Phonology</p> <p>Influence of Diachronic Language Processes on the Language Variability</p> <p>Diachronic Changes in Languages</p>
LEVEL	Introductory
MODALITY	Presentations and recorded presentations, accompanied by recommended additional reading and interspersed with interactive quizzes and other types of activities.
TIME COMMITMENT	80 hrs
ECTS	3

COURSE TITLE

INTRODUCTION TO TEXT PROCESSING AND ANALYSIS

COURSE INSTRUCTORS



PAVEL VONDŘÍČKA studied Nordic Studies (Norwegian and Icelandic) and German Studies at the Charles University in Prague and at the Icelandic University in Reykjavík. His interest in lexicography, lexical databases and corpus linguistics manifested itself in his doctoral dissertation where he could combine his earlier IT experience with the desire to promote innovations in the domain of humanities. At present, he keeps on developing tools for the construction and analysis of text corpora at the Institute of Czech National Corpus and helps promoting basic technical competences among students and researchers in philology.

Pavel.Vondricka@ff.cuni.cz



LUCIE CHLUMSKÁ graduated in Translation Studies (English) and Czech Studies at the Charles University in Prague, Czech Republic. After using language corpora in her Master's Thesis, she became fascinated by the potential of corpus linguistics and enrolled in a doctoral programme at the Institute of Czech National Corpus where she currently works. In her research, she is mainly interested in corpus-based translation studies and contrastive linguistics; she also helps promoting corpora in language teaching and learning, both L1 and L2.

lucie.chlumska@ff.cuni.cz

COURSE DESCRIPTION	<p>The main objective of the course is to provide beginner-level digital linguistics students with all the necessary information on text processing and analysis. Starting with basic topics, such as characteristics of a plain text format and the difference between data and metadata, the course goes on to explain the specifics of XML and different types of text annotation, to introduce the process of tokenization, segmentation and morphological analysis, to describe the limits and possibilities of syntactic and semantic tagging and, finally, to summarize the principles of CQL and corpus querying, including the use of regular expressions and querying parallel corpora.</p>
COURSE OBJECTIVES	<ul style="list-style-type: none"> • to understand how computers work with textual data • to distinguish between different data formats and extract textual content from them (e.g. using OCR) • to understand the specifics of plain text and XML formats • to understand the principles and issues of text annotation, incl. morphological analysis, syntactic and semantic tagging • to learn about available resources for text processing and analysis, including taggers and concordancers • to be able to analyse existing as well as own corpora in a variety of available corpus-based tools • to build complex CQL queries, including regular expressions and logical operators
TOPICS COVERED	<p>File formats related to textual data Plain text: Encoding, data and metadata Extensible Markup Language or XML Regular expressions Tokenization and corpus-data formats Morphological analysis: principles and tools Syntactic and semantic annotation Corpus exploration and analysis Querying corpus data with CQL Text alignment and parallel corpora</p>
LEVEL	Introductory
MODALITY	Interactive presentations, audio & screen recordings, exercises, knowledge quizzes, readings
TIME COMMITMENT	100 hrs
ECTS	4

COURSE TITLE

POST-EDITING MACHINE TRANSLATION

COURSE INSTRUCTORS



JEAN NITZKE studied at the Faculty of Translation Studies, Linguistics and Cultural Studies in Gernersheim, Johannes Gutenberg University Mainz, where she received her BA and MA degrees. Afterwards she worked briefly as an in-house translator and started her doctoral studies focusing on problem-solving in translation and post-editing. In 2012, she started a teaching and research position at the above-mentioned Faculty, which she still holds today. In November 2017, she defended her doctoral thesis and will now start to work on her post-doc project at the TRACO Center. Her research interests also include translation technologies, LSP translations, and cognitive translation studies.

nitzke@uni-mainz.de



ANKE TARDEL studied at the Faculty of Translation Studies, Linguistics and Cultural Studies in Gernersheim, where she received her BA and her MA degrees in Translation. Besides her studies, she has been a student research assistant since 2012, supporting various eye tracking studies and research projects.

As of October 2017, she is a PhD student, research assistant, and lecturer at the same faculty. Her research interests include translation process research, translation revision, post-editing, translation technologies, and cognitive translation studies.

antardel@uni-mainz.de

COURSE DESCRIPTION	This course will give you an introduction to post-editing (the correction of raw machine translated output by a human translator). It will cover both theoretical backgrounds and practical exercises.
COURSE OBJECTIVES	<ul style="list-style-type: none"> • to learn how machine translation works • to understand what post-editing is and how it can be executed • to distinguish between human translation, post-editing, and proofreading • to evaluate the factors influencing the post-editing process • to learn about post-editing in practice and research
TOPICS COVERED	<p>MT history</p> <p>MT approaches</p> <p>General PE</p> <p>Different PE styles</p> <p>PE and Translation Memory Systems</p> <p>PE and controlled languages</p> <p>PE in research</p> <p>PE in practice</p>
LEVEL	Introductory to intermediate
MODALITY	Interactive presentations, audio & screen recordings, exercises, knowledge quizzes, readings
TIME COMMITMENT	120 hrs
ECTS	5

COURSE TITLE

INTRODUCTION TO PYTHON FOR LINGUISTS

COURSE INSTRUCTORS



PETRA BAGO is Assistant Professor at the Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb. She is also Head of Natural Language Processing, Lexicography and Encyclopedic Science Section at the same Department. Her interests are digital humanities, e-lexicography, text encoding, natural language processing and statistical analysis. She teaches Python at the undergraduate and graduate level courses "Text and Language Processing", "Information Retrieval and Natural Language Processing", and "Natural Language Processing for Historical Texts".

petra.bago@gmail.com

COURSE DESCRIPTION	Python is one of the easiest programming languages out there right now. This course provides students with an understanding of elementary concepts in programming focusing on acquiring the knowledge and skills necessary for text processing. It is aimed at students of linguistics and other disciplines with no prior programming experience, who are interested in learning Python in order to process large volumes of text.
COURSE OBJECTIVES	<ul style="list-style-type: none"> • to understand and implement Python syntax and semantics • to identify, describe, and implement variables, operators and functions in Python • to identify, describe and implement integers, floating-point numbers, strings, lists, files and dictionaries in Python • to identify, describe and implement control flow in Python • to identify, describe and implement regular expressions in Python
TOPICS COVERED	<p>Introduction to Python</p> <p>Basic data types</p> <p>Variables</p> <p>Basic operators</p> <p>Basic functions</p> <p>Working with strings</p> <p>Working with lists</p> <p>Control flow</p> <p>Working with files</p> <p>Basic regular expressions</p> <p>Working with dictionaries</p>
LEVEL	Introductory
MODALITY	Interactive presentations, video & screen recordings, exercises, knowledge quizzes, readings, and assignments with instructor's feedback for student submissions.
TIME COMMITMENT	120 hrs
ECTS	5

COURSE TITLE

COMPUTATIONAL LEXICOLOGY AND LEXICOGRAPHY

COURSE INSTRUCTORS



NIVES MIKELIĆ PRERADOVIĆ is Associate Professor and Head of the Chair for Knowledge Organization at the Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb. Her work focuses on Natural Language Processing (language resources: valency lexicons, learner corpora), Service Learning (community-based learning and research), Multimedia Applications in Education and CALL. She has created the CroLTeC learner corpus.

nives.mikelic@gmail.com

COURSE DESCRIPTION

Computational lexicology may be defined as the application of computers to the study of the lexicon. Taken in its broadest sense, it is a multidisciplinary field involving the analysis of man-made dictionaries using computers to study their machine-readable text as well as a study of the computational linguistic content and organization of lexicons for use by natural-language processing applications. This course provides theoretical and practical information regarding current processes for building dictionaries and lexical databases used by natural-language processing applications. The topic is covered from the point of view of a computational lexicographer preparing dictionaries with the use of natural-language processing. Technical issues of dictionary building are also covered. In the project, students will explore dictionary entries in different computational lexicons that were built using the described tools, data and processes.

COURSE OBJECTIVES

- to understand the content and limitations of print dictionaries for computational purposes
- to critically compare the design, structure and content of various kinds of monolingual and bilingual subcategorisation (valency) lexicons.
- to explain the theoretical aspects and most important methods of building subcategorization lexicons
- to construct the valency entry in a bilingual valency lexicon
- to compare the design and content of various kinds of sentiment lexicons
- to plan a small-scale lexicographic project and implement it by applying the techniques discussed in class

TOPICS COVERED

Introduction to Computational lexicology and lexicography
Electronic lexicography, computational and corpus lexicography
Morphological lexicons
Derivational and inflectional morphological lexicons of different European languages
Lexical relations and lexical databases
Wordnets for different EU languages
Subcategorization (valency) lexicons
Semantic lexicons
Sentiment lexicons
Formats, standards and automatic acquisition of computational lexicons

LEVEL

Intermediate

MODALITY

Interactive presentations, video & screen recordings, exercises, knowledge quizzes, guided research tasks and assignments, directed readings

TIME COMMITMENT

120 hrs

ECTS

5

COURSE TITLE

LOCALIZATION TOOLS AND WORKFLOWS

COURSE INSTRUCTORS



CAROLINE REISS

As a graduate of the MA in Applied Translation Studies (2007), Caroline Reiss has been a freelance language service provider and Computer Assisted Translation (CAT) tools private trainer since 2006. She re-joined the Centre for Translation Studies, University of Leeds (UK) in 2012 and currently teaches modules in both Specialised Translation (Portuguese/Spanish into English) and CAT tools. She is interested in the impacts of technology and digital working environments on translators and in creating open-access learning resources for translator training.

C.L.Reiss@leeds.ac.uk

COURSE DESCRIPTION	Today's localization industry is faced with the task of translating a huge volume of texts to produce high-quality localized products in a short turnaround time, to satisfy the needs of global and local marketplaces. Such an undertaking would be inconceivable without the use of technology and, in recent decades, the development of localization tools has been instrumental and given rise to changes in collaborative workflows. This course aims to equip learners with a critical understanding towards the use of these tools, namely computer assisted translation (CAT) tools.
COURSE OBJECTIVES	<ul style="list-style-type: none"> • to describe the different types of computer tools available to support the localization process • to employ the basic functionalities of computer assisted translation tools • to combine the use of translation tools with other computer tools and resources to process language data • to assess the current limitations of these tools • to demonstrate an awareness of the level of planning and management involved in the localization process
TOPICS COVERED	<p>The historical development of the localisation industry</p> <p>The content production and localisation cycle</p> <p>Controlled language</p> <p>Translation technologies</p> <p>Basic concepts in translation memory software</p> <p>Localisation project lifecycle and management</p>
LEVEL	Introductory
MODALITY	Recorded presentations, interactive activities, quizzes, guided exploration and a project-based approach to learning about language technologies
TIME COMMITMENT	100 hrs
ECTS	4

COURSE TITLE

MINING AND MANAGING MULTILINGUAL TERMINOLOGY

COURSE INSTRUCTORS



SPELA VINTAR is Full Professor at the Department of Translation Studies, University of Ljubljana, where she lectures in Translation Technologies, Terminology Management, Machine Translation and Localisation. Her research interests involve various aspects of natural language processing and multilingual technologies, in particular multilingual terminology mining, knowledge modelling and MT evaluation. She is DigiLing's project leader and coordinator of an upcoming joint masters programme in Digital Linguistics.

spela.vintar@ff.uni-lj.si

COURSE DESCRIPTION	Systematic terminology management is essential for efficient communication of specialised content, either within the language industry, in business or institutional environments. This course provides students with an understanding of specialised discourse and an inventory of theoretical foundations, methods and tools for terminology management in multilingual contexts. A core component of this course requires students to develop the skills necessary for compiling domain-specific text collections and employing computational methods to extract specialised knowledge from such collections.
COURSE OBJECTIVES	<ul style="list-style-type: none"> • to recognise terms and their properties in different communicative settings • to understand the needs of various users of terminology • to plan, design and populate a termbase in any terminology management tool of choice • to compile a specialised corpus of texts to be used for terminology mining • to use a corpus workbench to extract single- and multi-word terms, definitions and other relevant information, in one or several languages • to validate, evaluate and implement the results of terminology mining
TOPICS COVERED	<p>Basic concepts in terminology</p> <p>Concepts, definitions and relations</p> <p>Terms across specialised domains, genres, languages</p> <p>Terminology management: structuring an entry, available tools</p> <p>Building a corpus for term mining</p> <p>Approaches to term mining</p> <p>Finding keywords and terms with a built-in term grammar</p> <p>Extracting terms and definitions with advanced CQL queries</p> <p>Multilingual term mining</p> <p>Validation and evaluation of results</p>
LEVEL	Intermediate
MODALITY	Interactive presentations, video & screen recordings, exercises, guided tasks and assignments, reading & self-guided research
TIME COMMITMENT	120 hrs
ECTS	5

