

MSC IN NLP

COURSE DESCRIPTION

SEMESTER 7 – S7

Updated for academic year 2024-2025

Study Unit – *Unité d'Enseignement (UE)*

Course – *Élément Constitutif (EC)*

● Mandatory

◆ ILPC Track

■ LI Track

▲ Optional

STUDY UNIT 701 – Probabilities, Statistics and Algorithms for AI

6 ECTS

COEF 6

● PROBABILITIES AND STATISTICS

Marianne CLAUSEL – marianne.clausel@univ-lorraine.fr

Mandatory for all M1 students.

This course revisits the fundamental mathematical properties of probability and statistics, covering classic data analysis methods and introducing more complex techniques such as SVMs and discriminant analysis, with practical applications using real data and R software.

Students will learn to:

- › Identify statistical analysis methods appropriate for studying datasets.
- › Perform statistical calculations for the mathematical analysis of datasets.
- › Conduct qualitative data evaluations.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

● PYTHON PROGRAMMING

Karën FORT – karen.fort@univ-lorraine.fr

Mandatory for all M1 students.

This course is designed to offer a comprehensive introduction to Python for beginners, and to focus on deepening Python skills to solve complex problems and develop high-level applications using various libraries for advanced learners.

Students will learn to:

- › Model a problem in the form of algorithms.
- › Develop a software solution in Python for data science problems.
- › Implement and apply classical algorithms for processing numerical data.
- › Master the Python programming environment.

CM – 15 hours

TD – 15 hours

TOTAL – 30 hours

COEF – 1

STUDY UNIT 702 – Written Corpora and Logic

6 ECTS

COEF 6

• WRITTEN CORPORA

Bruno GUILLAUME – bruno.guillaume@univ-lorraine.fr

Mandatory for all M1 students.

This course focuses on the methods and techniques used in NLP for creating and utilising written corpora, introducing corpus linguistics, criteria for corpus construction, data collection from the web, handling various document formats, and annotation procedures at multiple levels, along with the software for data exploitation.

Students will learn to:

- › Design the content of a written document corpus.
- › Standardise the data of a corpus.
- › Master the principles and examples of corpus annotation.

CM – 15 hours

TD – 15 hours

TOTAL – 30 hours

COEF – 1

• LOGIC

Philippe DE GROOTE – philippe.de-groote@univ-lorraine.fr

Mandatory for all M1 students.

This course aims to familiarise students with the fundamental concepts and methods of formal logic, presenting the syntax and semantics of propositional logic and predicate calculus, before addressing lambda calculus and higher-order logic.

Students will learn to:

- › Understand the syntax and semantics of classical logic.
- › Understand lambda calculus and higher-order logic.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

STUDY UNIT 703 – Tools and Challenges of NLP Development

6 ECTS

COEF 6

• ETHICS IN NLP

Karën FORT – karen.fort@univ-lorraine.fr

Mandatory for all M1 students.

This course introduces students to the ethical issues in NLP, including data privacy, algorithmic bias, information manipulation, and societal impact, through discussions and case studies that foster the ethical awareness necessary for designing responsible NLP systems and acting ethically in future AI-related work.

Students will learn to:

- › Master the development of solutions through an ethical approach.
- › Identify risk factors in the development of solutions.
- › Assess the consequences of developing an NLP solution.

CM – 7 hours

TD – 8 hours

TOTAL – 15 hours

COEF – 1

• PROJECT MANAGEMENT TOOLS

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Mandatory for all M1 students.

This course focuses on practical tools essential for effective project management, teaching students to use platforms like Git, work with the command line in Shell, collaborate on complex projects, and use LaTeX for documentation, enabling them to manage NLP projects professionally and collaboratively.

Students will learn to:

- › Design and develop an automatic data management system based on a real-world case.
- › Master techniques for gathering client requirements and modelling the corresponding functionalities.
- › Master project management techniques as a whole, including financial analysis (costs and time), group facilitation, and conflict management.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

• INTRODUCTION TO NLP

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Mandatory for all M1 students.

This course offers a comprehensive introduction to the fundamental concepts of NLP, while also exploring the relationship between NLP, computer science, and linguistics, and addressing the ecological impact of these advancements.

Students will learn to:

- › Understand the range of issues in NLP.
- › Use basic NLP tools on practical examples.
- › Implement a minimal solution to an NLP problem.

CM – 8 hours

TD – 7 hours

TOTAL – 15 hours

COEF – 1

STUDY UNIT 704 – Tracks (Specialisations)

6 ECTS

COEF 6

◆ **GENERAL LINGUISTICS**

Mathilde DARGNAT – mathilde.dargnat@univ-lorraine.fr

Mandatory for students who picked the "ILPC" track.

This course provides a comprehensive overview of general linguistics, introducing students to the foundational concepts and issues in the field. It covers a wide range of linguistic topics, from field linguistics to psycholinguistics, equipping students with essential theoretical knowledge and analytical skills for understanding language structure and use.

Students will learn to:

- › Understand the broad spectrum of linguistic challenges.
- › Use basic NLP tools on practical examples.

CM – 5 hours

TD – 5 hours

TOTAL – 10 hours

COEF – 1

◆ **PHONETICS**

Mélanie LANCIEN – melanie.lancien@univ-lorraine.fr

Mandatory for students who picked the "ILPC" track.

This course focuses on acoustic phonetics, targeting students with programming skills. It covers the basic principles of speech acoustics, followed by practical exercises in speech recognition, synthesis, and automatic processing of oral data, aiming to provide students with a solid foundation in the technical aspects of phonetics.

Students will learn to:

- › Implement analytical techniques for linguistic data.
- › Identify basic phonological properties.

CM – 5 hours

TD – 5 hours

TOTAL – 10 hours

COEF – 1

◆ **MORPHOLOGY**

Fiammetta NAMER – fiammetta.namer@univ-lorraine.fr

Mandatory for students who picked the "ILPC" track.

This course explores the study of morphology, focusing on the regular correlations between the form and meaning of complex words. Students will examine key concepts such as morphemes, lexemes, derivation, and inflection, as well as the construction of word forms in different languages.

Students will learn to:

- › Implement analytical techniques for linguistic data.
- › Identify the morphological properties of language.

CM – 5 hours

TD – 5 hours

TOTAL – 10 hours

COEF – 1

◆ SYNTAX

Marie-Laurence KNITTEL – marie-laurence.knittel@univ-lorraine.fr

Mandatory for students who picked the "ILPC" track.

This course delves into the structure of sentences and propositions. Students will learn how words are organised into grammatically correct phrases using key concepts such as noun phrases, verb phrases, subordination, coordination, and syntactic roles.

Students will learn to:

- › Approach sentence syntax as a formal, representable, and computable entity.

CM – 5 hours

TD – 5 hours

TOTAL – 10 hours

COEF – 1

◆ SEMANTICS

Sylvain POGODALLA – sylvain.pogodalla@univ-lorraine.fr

Mandatory for students who picked the "ILPC" track.

This course examines the meaning of words, phrases, and sentences. Students will explore semantic relationships such as synonymy, antonymy, metaphor, and metonymy, as well as theories of compositional semantics, enhancing their understanding of how meaning is constructed in language.

Students will learn to:

- › Understand and apply linguistic formalisms for semantics.

CM – 5 hours

TD – 5 hours

TOTAL – 10 hours

COEF – 1

■ FUNDAMENTAL COMPUTER SCIENCE

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Mandatory for students who picked the "LI" track.

This course, aimed towards students with a background in linguistics, explores essential computing skills for NLP and computational linguistics. Covering key concepts such as grammars, graphs, trees, automata, and regular expressions, students will learn to model and analyse the structure of natural languages, define production rules, represent hierarchical information, and solve NLP-related problems.

Students will learn to:

- › Understand fundamental concepts related to formal grammars.
- › Know how to represent and use graphs and trees.
- › Define automata and transducers.
- › Search for simple patterns in textual documents using regular expressions.

CM – 25 hours

TD – 25 hours

TOTAL – 50 hours

COEF – 1

STUDY UNIT 705 – Project and Language Courses

6 ECTS

COEF 6

• INTERDISCIPLINARY PROJECT

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Mandatory for all M1 students.

The project allows students to work on an open theme of their choice, guiding them in identifying the interdisciplinary skills needed for an NLP project while providing practical and conceptual project management tools.

Students will learn to:

- › Practise leading an interdisciplinary group project.
- › Undertake bibliographic research.
- › Manage a long-term project.
- › Implement a minimal solution to an NLP problem.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

• ENGLISH OR FRENCH (LANGUAGE COURSES)

José MENDES ANTUNES – jose.mendes@univ-lorraine.fr (French)

Carine MARTIN – carine.martin@univ-lorraine.fr (English)

Mandatory for all M1 students.

This course offers students a choice between scientific English or French, aiming to strengthen their language skills based on their profiles and needs within the Master's programme. Level-specific groups are provided.

Students will learn to:

- › Strengthen scientific English skills.

or

- › Master basic communication in French.

CM – 0

TD – 24 hours

TOTAL – 24 hours

COEF – 1

MSC IN NLP

COURSE DESCRIPTION

SEMESTER 8 – S8

Updated for academic year 2024-2025

Study Unit – *Unité d'Enseignement (UE)*

Course – *Élément Constitutif (EC)*

● Mandatory

◆ ILPC Track

■ LI Track

▲ Optional

STUDY UNIT 801 – Machine Learning and Symbolic AI

6 ECTS

COEF 6

● MACHINE LEARNING

Marianne CLAUSEL – marianne.clausel@univ-lorraine.fr

Mandatory for all M1 students.

This course introduces Bayesian models to cover fundamental principles of machine learning, including supervision, classification, and optimisation, while addressing industry scenarios. It emphasises practical aspects and best practices for utilising machine learning methods, such as data analysis, preprocessing, visualisation, quality measurement, and managing large and streaming corpora.

Students will learn to:

- › Apply statistical mathematical theories to real-world data produced by humans.
- › Analyse the behaviour of automated models to assess their quality.
- › Apply logical mathematical theories to internet-derived data.
- › Extract information from this data and structure it into reusable knowledge.

CM – 15 hours

TD – 15 hours

TOTAL – 30 hours

COEF – 1

● SYMBOLIC AI

Mathieu D'AQUIN – mathieu.d-aquin@univ-lorraine.fr

Mandatory for all M1 students.

The Symbolic AI course focuses on 'knowledge-centric' approaches to AI, i.e. problem-solving, decision support and learning methods based on the explicit and formalised representation of knowledge. Starting from traditional knowledge-based systems and associated knowledge engineering methodologies, the course focuses on methods and technologies associated with the semantic web, including knowledge graphs and ontologies.

Students will learn to:

- › Understand the associated challenges and approaches for representing and processing knowledge.
- › Implement symbolic approaches for specific AI use cases.

CM – 15 hours

TD – 15 hours

TOTAL – 30 hours

COEF – 1

STUDY UNIT 802 – Speech Corpora and Formal Tools

6 ECTS

COEF 6

• SPEECH CORPORA

Yves LAPRIE – y.laprie@univ-lorraine.fr

Mandatory for all M1 students.

This course covers speech modalities, including sound, facial images, and gestures, along with the technological devices for data collection and associated technical and ethical constraints.

Students will learn to:

- › Understand the specificities of spoken corpora.
- › Design the content of spoken corpora.
- › Annotate a spoken corpus.

CM – 15 hours

TD – 15 hours

TOTAL – 30 hours

COEF – 1

♦ FORMAL LANGUAGES

Philippe DE GROOTE – philippe.de-groote@univ-lorraine.fr

Mandatory for students who picked the "ILPC" track.

This course covers the theoretical foundations of formal grammars and languages, their relationships, and connections to recognising machines and algorithms.

Students will learn to:

- › Identify the language classes generated by formal grammars.
- › Manipulate finite state automata from both a practical and theoretical perspective.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

♦ CALCULABILITY AND COMPLEXITY

Romain PÉCHOUX – romain.pechoux@univ-lorraine.fr

Mandatory for students who picked the "ILPC" track.

This course covers computability and algorithmic complexity, focusing on Turing machines, non-computable functions, undecidable problems, complexity measures (space, time), and distinct complexity classes (P, PSPACE, NP, EXPTIME).

Students will learn to:

- › Model a problem with a Turing machine.
- › Identify the complexity of an algorithmic problem.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

STUDY UNIT 802 – Speech Corpora and Applied Linguistics

6 ECTS

COEF 6

● SPEECH CORPORA

Yves LAPRIE – y.laprie@univ-lorraine.fr

Mandatory for all M1 students.

This course covers speech modalities, including sound, facial images, and gestures, along with the technological devices for data collection and associated technical and ethical constraints.

Students will learn to:

- › Understand the specificities of spoken corpora.
- › Design the content of spoken corpora.
- › Annotate a spoken corpus.

CM – 15 hours

TD – 15 hours

TOTAL – 30 hours

COEF – 1

■ MORPHOPHONOLOGY

Stéphanie LIGNON – stephanie.lignon@univ-lorraine.fr

Mandatory for students who picked the "LI" track.

This course briefly reviews major phonological theories and then explores lexical construction from a formal perspective and studies specific cases to explain apparent semantic inconsistencies through morphophonological analysis.

Students will learn to:

- › Identify the constraints for proper formation of lexemes in French.
- › Understand the sense/shape interactions involved in morphological constructions.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

■ LEXICON–SYNTAX RELATION AND POLYSEMY

Alain POLGUÈRE – alain.polguere@univ-lorraine.fr

Mandatory for students who picked the "LI" track.

This course explores the semantic and syntactic properties of event-denoting nouns, both derived and non-derived. It focuses on how deverbal nouns inherit their syntactic and semantic features, including argument structure and aspectual structure, from their verb bases.

Students will learn to:

- › Identify the syntactic and semantic properties of various noun classes to assess their type and mode of reference.
- › Justify analyses based on specific evidence.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

STUDY UNIT 803 – Language Data Processing

6 ECTS

COEF 6

• DATA STORAGE AND RETRIEVAL

Anna NIKIFOROVSKAJA – anna.nikiforovskaja@univ-lorraine.fr

Mandatory for all M1 students.

This course familiarises students with techniques for retrieving and analysing data from various sources (Excel files, databases, web, tweets) and formats (XML, HTML, OWL, CSV, JSON).

Students will learn to:

- › Identify the tools and libraries needed to collect data.
- › Implement these tools and libraries in practical cases.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

• DATA ANALYSIS

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Mandatory for all M1 students.

This course covers data processing, analysis, and visualisation techniques. It explains data preprocessing and summarisation, predictive methods, and visualisation techniques. Additionally, it introduces Python libraries for statistical learning, such as pandas, scikit-learn, and SciPy.

Students will learn to:

- › Analyse data.
- › Pre-process and summarise data.
- › Visualise properties of data sets.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

● NEURAL NETWORKS

Claire GARDENT – claire.gardent@univ-lorraine.fr

Mandatory for all M1 students.

This course introduces neural networks, focusing on their use in machine learning and artificial intelligence. It covers basic concepts, common architectures, and tools for quick and easy implementation.

Students will learn to:

- › Design and implement neural networks, choosing the appropriate architectures for different tasks.
- › Train neural networks using optimisation techniques, and understand the mechanisms of machine learning.
- › Evaluate the performance of neural network models using appropriate metrics and compare them to other approaches.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

STUDY UNIT 804 – Linguistics

6 ECTS

COEF 6

● PROSODY

Mélanie LANCIEN – melanie.lancien@univ-lorraine.fr

Samantha RUVOLETTA – samantha.ruvoletto@univ-lorraine.fr

Mandatory for all M1 students.

This course examines prosody, focusing on melody, rhythm, and tonal variations in speech. Students will study how prosodic features affect meaning and expressiveness, and learn to identify variations across languages and contexts, with a focus on language acquisition and bilingualism.

Students will learn to:

- › Identify variations in pitch, rhythm, and melody in speech.
- › Analyse intonational and rhythmic patterns in speech.
- › Understand how prosody influences the meaning and interpretation of discourse.

CM – 6

TD – 6

TOTAL – 12

COEF – 1

● PHONETICS

Mélanie LANCIEN – melanie.lancien@univ-lorraine.fr

Mandatory for all M1 students.

This course explores the acoustic properties of speech sounds. Students will analyze dialects and communication variations, focusing on the material aspects of speech and their role in effective communication.

Students will learn to:

- › Analyse the phonological rules governing the pronunciation of words.
- › Identify phonological and dialectal variations.
- › Compare the phonological systems of different languages.

CM – 6

TD – 6

TOTAL – 12

COEF – 1

• MORPHOLOGY

Fiammetta NAMER – fiammetta.namer@univ-lorraine.fr

Mandatory for all M1 students.

This course examines the relationship between word form and meaning in complex words. It focuses on modern morphological principles, particularly constructional morphology, and explores irregular word formations such as templates, cross-formations, subtractive morphology, and parasynthesis, using relational and paradigmatic approaches to analyse these constructions.

Students will learn to:

- › Analyse the morphological structure of words by identifying morphemes and lexemes.
- › Understand the processes of word formation, including derivation and compounding.
- › Apply morphological skills to the analysis of morphology in different languages.

CM – 6

TD – 6

TOTAL – 12

COEF – 1

• SYNTAX

Alain POLGUÈRE – alain.polguere@univ-lorraine.fr

Mandatory for all M1 students.

This course covers sentence structure, with a descriptive and typological approach in the first part, followed by formal syntax within generative grammar in the second part, focusing on syntactic tree representations.

Students will learn to:

- › Understand the fundamental principles of sentence structure.
- › Analyse the syntactic rules that govern sentence structure.
- › Identify the dependency relations between elements of a sentence.

CM – 6

TD – 6

TOTAL – 12

COEF – 1

• SEMANTICS

Philippe DE GROOTE – philippe.de-groote@univ-lorraine.fr

Mandatory for all M1 students.

This course covers lexical and sentence semantics, focusing on relationships like synonymy, antonymy, and hyponymy, and teaches how language expresses and interprets meaning in communication.

Students will learn to:

- › Understand the lexical meaning of words and expressions.
- › Analyse the semantics of sentences and linguistic statements.
- › Use formal semantics tools to analyse meaning.

CM – 6

TD – 6

TOTAL – 12

COEF – 1

STUDY UNIT 805 – Project and Language Courses

6 ECTS

COEF 6

- **SUPERVISED PROJECT**

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Mandatory for all M1 students.

M1 students undertake a small group project to gain initial project management experience and familiarise themselves with laboratory teams and topics. Students are expected to work approximately half a day per week on subjects introduced in the previous semester.

Students will learn to:

- › Practice leading an interdisciplinary group project.
- › Prepare a written presentation of the results (introduction to article writing).
- › Prepare an oral presentation of the results (conference-style).

CM – Ø

TD – Ø

TOTAL – Ø

COEF – 1

- **ENGLISH OR FRENCH (LANGUAGE COURSES)**

José MENDES ANTUNES – jose.mendes@univ-lorraine.fr (French)

Carine MARTIN – carine.martin@univ-lorraine.fr (English)

Mandatory for all M1 students.

This course offers students a choice between scientific English or French, aiming to strengthen their language skills based on their profiles and needs within the Master's programme. Level-specific groups are provided.

Students will learn to:

- › Strengthen scientific English skills.

or

- › Master basic communication in French.

CM – 0

TD – 24 hours

TOTAL – 24 hours

COEF – 1

▲ INTERNSHIP

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Optional.

This allows students to undertake an internship in a laboratory or company between M1 and M2.

MSC IN NLP

COURSE DESCRIPTION

SEMESTER 9 – S9

Updated for academic year 2024-2025

Study Unit – *Unité d'Enseignement (UE)*

Course – *Élément Constitutif (EC)*

● Mandatory

◆ ILPC Track

■ LI Track

▲ Optional

STUDY UNIT 901 – Basics in NLP

9 ECTS

COEF 9

● NEURAL NETWORKS

Emmanuel VINCENT – emmanuel.vincent@univ-lorraine.fr

Mandatory for all M2 students.

This course provides an overview of neural network-based machine learning, covering topics from multi-layer perceptrons to advanced architectures like deterministic/probabilistic networks, convolutional networks, and recurrent networks.

Students will learn to:

- › Master theories on neural networks.
- › Apply these theories to develop a solution.

CM – 15 hours

TD – 15 hours

TOTAL – 30 hours

COEF – 1

● WRITTEN DATA PROCESSING

Kamal SMAILI – kamal.smaili@univ-lorraine.fr

Mandatory for all M2 students.

This course introduces learning methods and specialised libraries (e.g., NLTK, Giza) used for modelling various subtasks and applications in written text processing, with a focus on automatic translation and the Expectation-Maximisation algorithm.

Students will learn to:

- › Use NLP tools for standard tasks.
- › Implement techniques for machine translation.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

• LOW-RESOURCE LANGUAGES

Delphine BERNHARD – dbernhard@unistra.fr

Mandatory for all M2 students.

This course focuses on the study and preservation of low-resource languages, exploring challenges in maintaining linguistic diversity and revitalising endangered languages through sociolinguistic, educational, and technological strategies.

Students will learn to:

- › Understand the challenges of preserving and revitalising low-resource languages.
- › Master strategies for linguistic documentation and revitalisation.
- › Contribute to the preservation and revitalisation of low-resource languages through concrete actions.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

• GENERATIVE AI

Claire GARDENT – claire.gardent@univ-lorraine.fr

Mandatory for all M2 students.

This course explores text generation as a key element of large language models, focusing on generating content from data and text, including simplification and automatic summarisation, while introducing various symbolic, probabilistic, and statistical approaches.

Students will learn to:

- › Apply the principles of text generation.
- › Master the operation of automatic generation systems.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

• LEXICAL RESOURCES

Bruno GUILLAUME – bruno.guillaume@univ-lorraine.fr

Mandatory for all M2 students.

This course introduces key types of digital lexical resources, including electronic dictionaries, lexical databases, lexical ontologies, and transformer-based embeddings, grounded in foundational lexicology concepts.

Students will learn to:

- › Familiarise oneself with existing lexical resources.
- › Build specialised lexicons.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

• DIALOGUE, CHATBOTS AND QA SYSTEMS

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Mandatory for all M2 students.

This course explores recent advancements in human-machine interaction and covers fundamental design principles, development methods, and underlying technical aspects.

Students will learn to:

- › Design, develop, and evaluate intelligent dialogue systems for various applications, ranging from customer support to human-robot interaction.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

STUDY UNIT 902 – Openings

9 ECTS

COEF 9

• SYNTACTIC MODELS

Sylvain POGODALLA – sylvain.pogodalla@univ-lorraine.fr

Mandatory for all M2 students.

This course presents fundamental concepts of symbolic grammatical formalisms for describing syntax and the syntax-semantics interface in natural language, focusing on various formal systems such as categorical grammars, tree adjoining grammars, and lexical functional grammars.

Students will learn to:

- › Analyse a set of linguistic examples to derive generalisations.
- › Implement the modelling of a linguistic phenomenon using an existing formalism.
- › Identify the flaws and limitations of a formalisation.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

• SEMANTICS AND DISCOURSE

Philippe DE GROOTE – philippe.de-groote@univ-lorraine.fr

Mandatory for all M2 students.

This course introduces various methods for representing the meaning of complex expressions at the sentence level, focusing on lambda calculus and higher-order logic, and exploring how the meaning of complex expressions is constructed from their components, as well as phenomena in discourse that require contextual and hierarchical considerations.

Students will learn to:

- › Understand theories of semantic and discourse representation.
- › Apply lambda calculus and logic to semantics.

- › Apply discourse representation theories.
- › Solve linguistic problems at the discourse level.

CM – 15 hours

TD – 15 hours

TOTAL – 30 hours

COEF – 1

• SPEECH RECOGNITION AND SYNTHESIS

Emmanuel VINCENT – emmanuel.vincent@univ-lorraine.fr

Mandatory for all M2 students.

This course explores methods for converting speech to text and generating natural speech from text, while addressing the principles of automatic speech recognition and the design of effective speech synthesis systems for various applications.

Students will learn to:

- › Design and evaluate speech recognition and synthesis systems.
- › Use the technologies required for voice technology, accessibility, and conversational artificial intelligence.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

• SIGNAL PROCESSING

Fateme GHAYEM – fatemeh.ghayem@inria.fr

Vincent MARTIN – vincent.martin@inria.fr

Mandatory for all M2 students.

This course teaches the theoretical foundations and practical applications of audio signal techniques, including acquisition, preprocessing, analysis, and advanced topics like Fourier transforms and adaptive filtering for applications such as speech recognition and audio pattern detection.

Students will learn to:

- › Apply signal processing techniques in various contexts.
- › Solve real-world signal processing problems.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

• ONTOLOGY

Mathieu D'AQUIN – mathieu.d-aquin@univ-lorraine.fr

Mandatory for all M2 students.

This course examines the relationship between the lexical and ontological levels, focusing on the processing of poly-lexical terms in specialised fields, utilising TTC TermSuite software for analysis and classification, and exploring terminological resources and their connection to Linked Open Data.

Students will learn to:

- › Build domain ontologies.
- › Manipulate terminological resources.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

• INTELLIGENT SYSTEMS

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Mandatory for all M2 students.

This course focuses on machine learning aspects of recommendation systems, particularly those based on usage data from social networks, guiding students in evaluating these systems through case studies and providing a software toolkit for independent experimentation.

Students will learn to:

- › Identify usage traces of a system.
- › Understand recommendation mechanisms.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

STUDY UNIT 903 – “ILPC” Track

6 ECTS

COEF 6

◆ LARGE LANGUAGE MODELS

Christophe CERISARA – christophe.cerisara@loria.fr

Mandatory for students who picked the “ILPC” track.

This course immerses students in large language models, exploring their architecture, training mechanisms, and practical applications for generating grammatically correct and semantically relevant text, while fostering critical evaluation skills for measuring model effectiveness in various contexts.

Students will learn to:

- › Understand the basic principles of designing large language models.
- › Apply these models to generate coherent and relevant text.
- › Evaluate the performance of language models using appropriate metrics for various natural language processing tasks.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

◆ SYMBOLIC KNOWLEDGE DISCOVERY

Miguel COUCEIRO – miguel.jorge-couceiro-de-sousa-santos@univ-lorraine.fr

Mandatory for students who picked the "ILPC" track.

This course introduces students to advanced techniques for symbolic knowledge extraction from unstructured textual data.

Students will learn to:

- › Use techniques for extracting concepts and relationships from unstructured texts.
- › Create symbolic knowledge bases from textual data.
- › Integrate these knowledge bases into AI systems to enhance understanding and decision-making.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

◆ REASONING IN AI

Zied BOURAOU – zied.bouraoui@cril.fr

Mandatory for students who picked the "ILPC" track.

This course explores the philosophy and concepts of symbolic artificial intelligence, focusing on symbolic representations, reasoning rules, and cognitive modelling, while analysing the strengths and limitations of symbolic AI and hybrid approaches, alongside ethical questions related to these systems.

Students will learn to:

- › Analyse the philosophical foundations of symbolic AI and its implications for modelling artificial intelligence.
- › Discuss the limitations and advantages of the symbolic approach compared to deep learning-based methods.
- › Evaluate the ethical and societal implications of symbolic AI in the context of technological evolution.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

◆ ROBUST SPEECH PROCESSING

Romain SÉRIZEL – romain.serizel@univ-lorraine.fr

Mandatory for students who picked the "ILPC" track.

This course focuses on the robustness of speech recognition systems, addressing variability in speech, regional accents, background noise, and acoustic distortions, while teaching students to identify vulnerabilities and enhance system reliability in real-world situations through practical case studies.

Students will learn to:

- › Identify the specific challenges related to speech variability and sources of acoustic disturbance.
- › Develop techniques to improve the resilience of speech recognition systems under adverse conditions.

- › Evaluate the performance of speech recognition systems, considering robustness and reliability in various contexts.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

◆ OPINION ANALYSIS AND MULTILINGUALISM

Margot MIESKES – margot.mieskes@h-da.de

Mandatory for students who picked the "ILPC" track.

This course focuses on opinion analysis and multilingualism by processing textual data in various languages, exploring techniques such as comparable corpora, document alignment methods, and teaching students to detect opinions using advanced learning algorithms.

Students will learn to:

- › Collect and process textual data in multiple languages for opinion analysis tasks.
- › Build sentiment analysis models capable of handling multilingual texts.
- › Interpret the results of opinion analysis to extract useful and relevant information in a multicultural and multilingual context.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

STUDY UNIT 903 – "LI" Track

6 ECTS

COEF 6

■ PROMPT ENGINEERING

Claire GARDENT – claire.gardent@univ-lorraine.fr

Mandatory for students who picked the "LI" track.

This course presents the technical programming method for prompt generation. The students will apply these methods to generate different complexity of prompts for the main LLMs.

Students will learn to:

- › Develop the ability to craft effective prompts through experience and intuition.
- › Analyse and refine prompts for optimal performance.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

■ SPEECH PROCESSING

Yves LAPRIE – y.laprie@univ-lorraine.fr

Mandatory for students who picked the "LI" track.

This course explores the link between phonetics, phonology, and automatic speech processing, covering advanced techniques in speech recognition, synthesis, prosody analysis,

and addressing challenges like pronunciation variability, regional accents, and background noise management.

Students will learn to:

- › Understand the basic principles of speech recognition and speech analysis.
- › Apply advanced techniques for analysing the acoustic and prosodic features of speech.
- › Develop speech processing systems for applications such as automatic transcription or speech synthesis.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

■ PRAGMATICS

Mathilde DARGNAT – mathilde.dargnat@univ-lorraine.fr

Mandatory for students who picked the "L1" track.

This course explores linguistic pragmatics, focusing on effective communication through concepts like reference, presupposition, implicature, and speech acts, while connecting these to formal representations in semantics and discourse theory.

Students will learn to:

- › Analyse speech acts and implicatures in real communicative contexts.
- › Understand how pragmatics influences the meaning of statements beyond syntax.
- › Apply pragmatic principles to analyse discourse and complex linguistic interactions.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

■ LEXICOLOGY

Yvon KEROMNES – yvon.keromnes@univ-lorraine.fr

Mandatory for students who picked the "L1" track.

This course introduces lexicology and lexicography, focusing on synchronic and diachronic analysis, differences between them, and translating lexicological research into dictionary descriptions.

Students will learn to:

- › Study the structure and formation of words in different languages.
- › Analyse the semantic properties of lexemes and their historical evolution.
- › Use lexical analysis methods to extract relevant lexical information from texts.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

■ TERMINOLOGY

Alain POLGUÈRE – alain.polguere@univ-lorraine.fr

Mandatory for students who picked the "LI" track.

This course focuses on terminology, teaching students to identify, extract, and standardise domain-specific terms, while exploring methods for building terminological databases and aligning multilingual terminologies.

Students will learn to:

- › Create glossaries and specialised terminology databases for specific domains.
- › Manage multilingual terminologies and align technical terms across different languages.
- › Apply terminological skills for specialised translation and information retrieval in professional contexts.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

■ DIALOGUE ENGINEERING

Maxime AMBLARD – maxime.amblard@univ-lorraine.fr

Mandatory for students who picked the "LI" track.

This course presents the definitions of linguistic models of dialogue, construction of a corpus, definition of an annotation system and production of a resource usable for chatbots.

Students will learn to:

- › Design interactive dialogue systems tailored to user needs.
- › Develop chatbots capable of engaging in natural and useful conversations.
- › Implement automated question-answering systems to effectively respond to a variety of questions in natural language.

CM – 7 hours

TD – 7 hours

TOTAL – 14 hours

COEF – 1

STUDY UNIT 904 – Project and Language Courses

6 ECTS

COEF 6

● SOFTWARE PROJECT

Karën FORT – karen.fort@univ-lorraine.fr

Mandatory for all M2 students.

This course focuses on developing essential programming and software engineering skills, teaching students to design, develop, and test software through practical projects while familiarising them with Agile methodologies and best practices in software quality.

Students will learn to:

- › Master programming techniques and be able to design functional software.
- › Apply software project management tools to plan and execute development projects.
- › Evaluate software quality by testing and debugging code effectively.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

• **ENGLISH OR FRENCH (LANGUAGE COURSES)**

José MENDES ANTUNES – jose.mendes@univ-lorraine.fr (French)

Carine MARTIN – carine.martin@univ-lorraine.fr (English)

Mandatory for all M2 students.

This course offers students a choice between scientific English or French, aiming to strengthen their language skills based on their profiles and needs within the Master's programme. Level-specific groups are provided.

Students will learn to:

- › Strengthen scientific English skills.

or

- › Master basic communication in French.

CM – 0

TD – 24 hours

TOTAL – 24 hours

COEF – 1

• **ETHICS AND ORIENTATION**

Karën FORT – karen.fort@univ-lorraine.fr

Mandatory for all M2 students.

This course trains students to respect participants' rights regarding personal data collection, storage, and publication, while exploring ethical dimensions in academic and corporate research, addressing dilemmas related to data privacy, cybersecurity, and artificial intelligence.

Students will learn to:

- › Analyse ethical dilemmas related to technology and be able to make informed ethical decisions in professional situations.
- › Understand the various ethical issues regarding the creation, use, and dissemination of corpora.
- › Master the professional research and business environment.

CM – 10 hours

TD – 10 hours

TOTAL – 20 hours

COEF – 1

MSC IN NLP

COURSE DESCRIPTION

SEMESTER 10 – S10

Updated for academic year 2024-2025

Study Unit – *Unité d'Enseignement (UE)*

Course – *Élément Constitutif (EC)*

● Mandatory

◆ ILPC Track

■ LI Track

▲ Optional

STUDY UNIT 1001 – Internship

6 ECTS

COEF 6

● **INTERNSHIP**

Karën FORT – karen.fort@univ-lorraine.fr

Mandatory for all M2 students.

The second semester of the M2 is dedicated to an internship, either in a company or a laboratory, culminating in a thesis presentation before a jury.