

# LISBON SUMMER SCHOOL IN LINGUISTICS

## PhD Programme

**Lisboa, July 1-5, 2024**

Venue  
Faculdade de Ciências Sociais e Humanas – NOVA  
Avenida de Berna, 26-C  
1069-061 Lisboa – Portugal

### AREA 2: TERMINOLOGY AND LEXICOGRAPHY

---

**Course 1: Lexical relations: theory and practice** – Alain Polguère (Université de Lorraine, CNRS, ATILF - Nancy, France)

---

#### COURSE DESCRIPTION

The universal system of standard lexical functions put forward by Meaning-Text linguistics (Mel'čuk 1996) is a powerful theoretical and descriptive tool used to account for lexical choices performed by speakers in the process of sentence synthesis. It also allows for the modeling of paradigmatic and syntagmatic lexical relations that structure the lexicon of natural languages. This one-week course will introduce lexical functions in five steps:

1. preliminary lexicological and grammatical notions;
2. nature and role of lexical functions;
3. paradigmatic lexical functions;
4. syntagmatic lexical functions;
5. case study in the lexical field of 'feelings'.

The teaching will be based on the manuscript of a new account of lexical functions – Mel'čuk & Polguère (in preparation), which will be made available to course participants.

The course will be taught in ENGLISH.

#### References

- MEL'ČUK Igor (1996). Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In Leo Wanner (ed.): *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia, John Benjamins.
- MEL'ČUK Igor, POLGUÈRE Alain (in preparation). *Lexical Functions: The Naked Truth*.

#### Bio

Alain Polguère

<https://perso.atilf.fr/apolguere/fr/accueil/>

---

**Course 2: TEI Lex-0 in action: Dictionaries Modelling** – Ana Salgado (NOVA CLUNL | NOVA FCSH, Portugal) - *Limited to 15 attendees*

---

### **COURSE DESCRIPTION**

In the field of lexicography, global mass digitization projects have played a vital role in preserving heritage lexicographic collections and facilitating access to these historical documents. Simple digitization has become obsolete; it is now crucial to make printed dictionaries available on the web by converting them into computationally explorable and interoperable lexical resources, which are ensured through consistent modelling and encoding of lexicographic data.

This introductory course serves as a gateway to the world of dictionary encoding, following the Guidelines of the Text Encoding Initiative (TEI), with a special emphasis on TEI Lex-0. While the remarkable flexibility of TEI guidelines offers advantages in various circumstances, it is essential to recognize that they also present specific challenges in the context of interoperability. TEI Lex-0, as a customized version of the TEI schema, has been meticulously developed to ensure the desired interoperability among different lexicographic resources.

From theory to practice, the course provides a comprehensive overview of text encoding, exploring the fundamentals of TEI and key concepts such as encoding versus modelling, markup languages, metadata, formats, and standards. Another component of the course focuses on macro and microstructural analysis of printed dictionaries, including the identification of different lexicographic components of entries, as well as lexicographic conventions, essential topics for ensuring consistent modelling.

In the practical sessions, students will have the opportunity to analyze concrete examples extracted and resulting from the encoding of the MORDigital project (DOI 10.54499/PTDC/LLT-LIN/6841/2020) and work directly with real data, particularly on heritage dictionaries from the lexicographic archive of the Lisbon Academy of Sciences. By the end of the course, participants are expected to be able to identify dictionary structures and components and to have acquired basic skills in TEI Lex-0, enabling them to model dictionaries in a flexible and adaptable manner. Whether an experienced lexicographer or a novice, this course offers a unique opportunity to enhance skills and contribute to the advancement of academic projects through standardized dictionary encoding.

The course will be taught in PORTUGUESE and ENGLISH.

### **Bio**

Ana Salgado

<https://www.cienciavitae.pt/portal/291E-340D-2DE4>

---

**Course 3: Linguistic Linked Open Data for humanists** – Anas Fahad Khan (CNR-ILC, Italy) & Giulia Pedonese (CNR-ILC, Italy) & Michele Mallia (CNR-ILC, Italy)

---

## COURSE DESCRIPTION

Having achieved popularity as a way of publishing and accessing data in different fields of the sciences and for sharing large encyclopaedic datasets such as DBpedia (derived from Wikipedia), linked data is becoming more and more popular in different areas of the humanities. In this course we will present a comprehensive introduction to the creation, publication, and use of linked open data for anyone who wants to work with linguistic datasets -- such as lexicons and corpora -- and especially for those who come from a linguistic or humanist background. We will look at the basics of linked data and the Semantic Web and introduce the various different standards technologies that make up the Semantic Web stack before focusing on the particular case of linked data language resources. During the course we will study the most important tools, vocabularies, and resources available in the Semantic Web and provide hands-on training for the creation and querying of linguistic linked data. We will look at how Semantic Web technologies can contribute to the creation of FAIR language resources as well as how to publish your resource on the linked open data cloud. We will also show how the Semantic Web query language SPARQL can be a powerful tool for data exploration.

The course will be taught in ENGLISH.

## Bio

### Anas Fahad Khan

Researcher at CNR-ILC. He is part of the Language Resources and Infrastructures group at the Istituto di Linguistica Computazionale “A. Zampolli”. His areas of interest include lexical standards, linguistic linked data, and the overlap between digital humanities and ontologies. In particular he is interested in the retrodigitisation of legacy lexicographic resources, the creation of language resources for historical languages, and the conceptual modelling of diachronic lexical data. He has been nominated as an expert by the British Standards Institute and is an active member of the ISO working groups ISO/TC 37/SC 4/WG 4 and ISO/TC 37/SC 2/WG 9. He is also a chair of the W3C Ontology Lexicon Group and is also a member of both the CLARIN Standards and Interoperability Committee and the DARIAH Lexical Resources group.

### Giulia Pedonese

Is a Research Technologist at CNR-ILC working on the design of FAIR training materials to enhance the knowledge of CLARIN linguistic resources within the H2IOSC project. She obtained a PhD in Italian Studies at the University of Pisa with a thesis on Dante Alighieri’s language; she was a postdoctoral researcher at the Catholic University of the Sacred Heart of Milan where she contributed to the ERC project Lila-Linking Latin working on refining and extending the Latin Wordnet and the Latin Valency Lexicon; she was also involved in the linguistic annotation of other resources, such as UDante. The lexicographic data resulting from her PhD thesis were published as a LLOD resource

called Dante Latin Loanwords and linked to the LiLa Knowledge Base. She contributed for five years to the lexicographic project called “Vocabolario Dantesco Latino” developed with the Crusca Academy. Her research interests lie in Language Resources, Digital Lexicography and Linguistic Annotation.

### Michele Mallia

Research Technologist at the Institute of Computational Linguistics "A. Zampolli", and works in the intersection of computational linguistics and Linked Open Data (LOD). With a Master's Degree in Digital Humanities from Pisa, Michele has a strong background in the creation and management of linguistic resources; a concentration on ontologies, metadata, and semantic knowledge representation formats, notably OntoLex Lemon, has also been a defining aspect of their work. They participated in the PRIN 2017 Italia Antica project as a research fellow with a focus on the development of web applications for editing OntoLex Lemon resources and linking epigraphic materials to LOD resources for proto-Italic languages. Currently participating in the Italian National H2IOSC EU NextGeneration project, Michele is contributing to the creation of a federated national research infrastructure through the development of web applications for querying and managing LLOD resources. This experience serves to underscore a commitment to advancing the field of Linguistic Linked Open Data through innovative technological solutions.

---

## SCHEDULE

---

	July 1-5	Monday	Tuesday	Wednesday	Thursday	Friday
Area 1 courses	9:30am – 12:30pm	A.Castelo & C.Zhou	A.Castelo & C.Zhou	A.Castelo & C.Zhou	A.Castelo & C.Zhou	A.Castelo & C.Zhou
	2:00pm – 5:00pm	Cristina Flores	Cristina Flores	Cristina Flores	Cristina Flores	Cristina Flores
	5:30pm – 8:30pm	A.Fiéis et al.	A.Fiéis et al.	A.Fiéis et al.	A.Fiéis et al.	A.Fiéis et al.
Area 2 courses	9:30am – 12:30pm	Alain Polguère	Alain Polguère	Alain Polguère	Alain Polguère	Alain Polguère
	2:00pm – 5:00pm	Ana Salgado	Ana Salgado	Ana Salgado	Ana Salgado	Ana Salgado
	5:30pm – 8:30pm	F.Khan & G.Pedonese & M.Mallia	F.Khan & G.Pedonese & M.Mallia	F.Khan & G.Pedonese & M.Mallia	F.Khan & G.Pedonese & M.Mallia	F.Khan & G.Pedonese & M.Mallia
Area 3 courses	9:30am – 12:30pm	Célia Lopes	Célia Lopes	Célia Lopes	Célia Lopes	Célia Lopes
	2:00pm – 5:00pm	Laura Calabrese	Laura Calabrese	Laura Calabrese	Laura Calabrese	Laura Calabrese
	5:30pm – 8:30pm	C.Pinto & A.Guilherme	C.Pinto & A.Guilherme	C.Pinto & A.Guilherme	C.Pinto & A.Guilherme	C.Pinto & A.Guilherme

## FEES

### Fees for PhD Students at NOVA FCSH

- Enrolled in Doctoral Programs in Linguistics, Translation and Terminology, or Languages Teaching – Multilingualism and Education for a Global Citizenship: FREE
- Enrolled in any Doctoral Program at Universidade NOVA de Lisboa: 1st course. FREE; other courses: 50€ each

### Fees for all

- Registration in a single course: 90€
- Registration in 2 courses: 120€
- Registration in 3 courses: 150€

**NOTE:** Special payment conditions for participants covered by protocols with NOVA FCSH (to be defined on a case-by-case basis, depending on the agreement).

## REGISTRATION

Enrolment is open until **21<sup>st</sup> of June, 2024**.

To enrol, please fill in the form available [here](#) and consider the following payment information.

**Please, be aware that enrolment will only be validated after you send us the confirmation of transaction.**

## PAYMENT

Payment can be made by one of the following ways:

### Option 1 – Bank transfer:

IBAN: PT50 0018 000321419114020 13

Banco Santander Totta S.A.

BIC/SWIFT: TOTAPTPL

Holder: FCSH – Univ. Nova de Lisboa

CIF / VAT number: 501559094

Address: Av. Berna 26 C, Lisboa 1069-061

**Option 2 – Paypal:**

Send payment to the account: [dgfc@fcsch.unl.pt](mailto:dgfc@fcsch.unl.pt)

In the transfer details, you must specify “Summer School – NOVA CLUNL”.

**Afterwards, please send the confirmation of transaction to [sec.clunl@fcsch.unl.pt](mailto:sec.clunl@fcsch.unl.pt)**

**ACCOMMODATION**

You can find a list of nearby hotels [here](#).