



ELEXIS: Training and Education

Toma Tasovac

Belgrade Center for Digital Humanities



Why does ELEXIS need training and education?

- ELEXIS: knowledge network and service provider
- Diverse user base
- Training and education is a core requirement for the social sustainability of research infrastructures

WP5 Objectives

- evaluate the skills researchers need in order to benefit from and contribute to the lexicographic infrastructure.
- develop online training materials covering the full spectrum of producing, exploiting and disseminating lexicographic data.
- develop and enhance online tutorials for a variety of ELEXIS tools and services.
- organize and conduct ELEXIS training measures

Deliverables

- **D5.1 ELEXIS Skillset Report (M12)** Primary data for this report was gathered via qualitative interviews with lexicographers, service providers, ELEXIS working groups, DARIAH and CLARIN.
- **D5.2 Guidelines for Producing ELEXIS Tutorials and Instruction Manuals (M24)** A set of practical recommendations for developing tutorials and instruction manuals for ELEXIS tools and services that lower the barrier for participation in the infrastructure.
- **D5.3 Overview of Online Tutorials and Instruction Manuals (M42)** A clear, structured overview of tutorials and instruction manuals developed within the project will be made available on the ELEXIS website.

Deliverables (translated into non-EU speak)

- D5.1 was about what we need to teach
- D5.2 was be about how and where
- Y3 and Y4 were about producing ELEXIS Learning Resources (and we're still there)

Task 5.1 Skills Assessment - Key Findings

- Almost all of our informants received no formal lexicographic training
- Almost all of our informants described themselves as lexicographic self-learners who honed their skills by:
 - consulting various dictionaries;
 - reading the scholarly literature;
 - attending workshops and seminars outside the regular university curricula; and/or
 - learning from their colleagues or on the job.

Task 5.1 Skills Assessment - Key Findings

- Lack of university curricula in lexicography or systemic, integrated training outside the university
 - *The knowledge gained in unconnected workshops is partial and segmented* (Hudeček)
- If lexicography is taught as a university subject, it is in some countries too theoretical and removed from contemporary trends and best practices
- If students have a background in linguistics, they don't necessarily have the necessary technical skills, and vice versa

Task 5.1 Skills Assessment - Key Findings

- **Technology is not the ultimate challenge: the reluctance to adopt new methods is also a social and cultural challenge**
- **Not all lexicographers believe in a culture of open access, open source and sharing lexicographic data.**



Task 5.1 Skills Assessment - Key Findings

- We should support critical use of tools
 - *The tools of today are not necessarily the tools of tomorrow. We should prepare students not only how to learn tools but also to be able to switch to different tools.* (Costa)
- Training materials should be on a wide range of topics and openly accessible to everybody
 - *ELEXIS should contribute towards setting a general training standard as well as a reference point for lexicographic practice in general.* (Grønvik)
 - *ELEXIS could play an important role in integrated curriculum development.* (Hudeček)

Task 5.2 Guidelines – Toward an ELEXIS Curriculum

- Based on the Skillset Report, Survey on User Needs and internal consultations, we decided to work toward an **ELEXIS Curriculum**
 - to embed the description of our services into a broader training narrative (countering the lack of university curricula or systemic, integrated training opportunities outside the university)
 - to provide a healthy balance between theory and practice cover the basics (“Do not assume that anybody knows anything. Think how low you can go.”)
 - to create a “specific website” or a “platform”



Task 5.2 ELEXIS Curriculum (intro, intermediate and advanced)

- Introduction to Dictionaries
- Introduction to Dictionary Users
- Introduction to Corpus-Based Lexicographic Practice
- Capturing, Modeling and Transforming Lexical Data: An Introduction
- LEX2 Mastering ELEXIS Corpus Tools for Lexicographic Purposes
- Lexonomy: Mastering the ELEXIS Dictionary Writing System
- Automating the Process of Dictionary Creation
- CLARIN Tools and Resources for Lexicographic Work
- Lexicography in the Age of Open Data
- Standards for Representing Lexical Data: An Overview
- Modeling Dictionaries in TEI Lex-0
- Modeling Dictionaries in OntoLex-Lemon
- LEX3: Transforming Legacy Dictionaries using Elexifier
- LEX3: Publishing Legacy Dictionaries with Publex
- Mastering LEX1: The Dictionary Matrix
- Mastering oXygen XML Editor for Dictionary Nerds
- Extracting Lexical Data: XPath for Dictionary Nerds
- Transforming Lexical Data: XSLT for Dictionary Nerds
- Trends in Computational Lexicography: A Pathfinder for Developers

Task 5.2 ELEXIS Curriculum (Contextualization, ELEXIS Tools and Other Tools)

- Introduction to Dictionaries
- Introduction to Dictionary Users
- Introduction to Corpus-Based Lexicographic Practice
- Capturing, Modeling and Transforming Lexical Data: An Introduction
- LEX2 Mastering ELEXIS Corpus Tools for Lexicographic Purposes
- Lexonomy: Mastering the ELEXIS Dictionary Writing System
- Automating the Process of Dictionary Creation
- CLARIN Tools and Resources for Lexicographic Work
- Lexicography in the Age of Open Data
- Standards for Representing Lexical Data: An Overview
- Modeling Dictionaries in TEI Lex-0
- Modeling Dictionaries in OntoLex-Lemon
- LEX3: Transforming Legacy Dictionaries using Elexifier
- LEX3: Publishing Legacy Dictionaries with Publex
- Mastering LEX1: The Dictionary Matrix
- Mastering oxygen XML Editor for Dictionary Nerds
- Extracting Lexical Data: XPath for Dictionary Nerds
- Transforming Lexical Data: XSLT for Dictionary Nerds
- Trends in Computational Lexicography: A Pathfinder for Developers

DARIAH-Campus

- central hub for DH training and education
- both a discovery layer (for remote resources) and a hosting platform
- Harvested by the SSHOC (Social Sciences and Humanities Open Cloud) Marketplace
- Onboarded on EOSC (European Open Science Cloud)
- Stability, versioning, sustainability



The ELEXIS Curriculum

AUTHORS



Toma Tasovac

TOPICS

LEXICOGRAPHY

CURRICULUM OVERVIEW

Introduction to Dictionaries

Introduction to Dictionary Users

Introduction to Corpus-Based
Lexicographic Practice

Capturing, Modeling and
Transforming Lexical Data: An
Introduction

LEX2: Mastering ELEXIS Corpus
Tools for Lexicographic Purposes

Lexonomy: Mastering the ELEXIS
Dictionary Writing System

Automating the Process of
Dictionary Creation

CLARIN Tools and Resources for
Lexicographic Work

Standards for Representing
Lexicographic Data: An Overview

Modeling Dictionaries in TEI Lex-0

Modeling Dictionaries in OntoLex-
Lemon

The ELEXIS Curriculum is an integrated set of training materials which contextualizes ELEXIS tools and services inside a broader, systematic pedagogic narrative. This means that the goal of the ELEXIS Curriculum is not simply to inform users about the functionalities of particular tools and services developed within the project, but to show how such tools and services are a) embedded in both lexicographic theory and practice; and b) representative of and contributing to the development of digital skills among lexicographers.

Resources

Introduction to Dictionaries

The goal of this course is to introduce a brief history of dictionaries as tools for the organization of knowledge about words and their meanings, and to analyze different ways of understanding and classifying the dictionary genre.

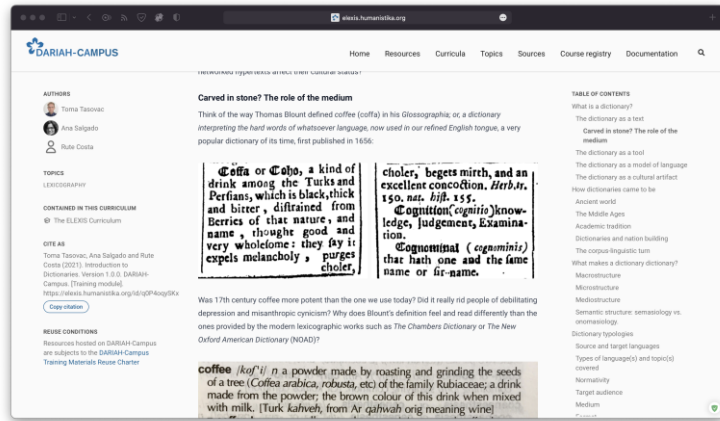


[Read more →](#)

Not all roads lead to Rome (or Florence, for that matter)

- Different paths through the curriculum
- Different learning scenarios

From zero to hero...



Carved in stone? The role of the medium

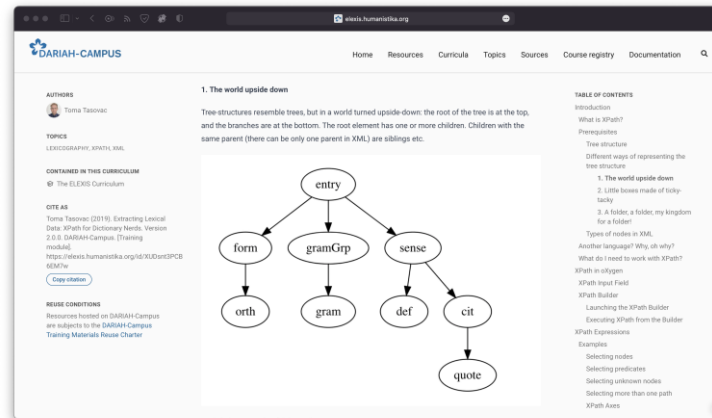
Think of the way Thomas Blount defined coffee (coffa) in his Glossographic, or a dictionary interpreting the hard words of whatsoever language, now used in our refined English tongue, a very popular dictionary of its time, first published in 1656:

Coffa or Coffo, a kind of drink among the Turks and Perfians, which is black, thick and bitter, diffrained from Berries of that nature, and name, thought good and very wholesome: they fay it expels melancholy, purges cholier, begets mirth, and an excellent concoction. Herb. n. 150. n. s. h. 155. Cognation, cognitie) knowledge, Judgement, Examination. Cognitional (cognizant) that hath one and the same name or fir-name.

Was 17th century coffee more potent than the one we use today? Did it really nd people of debilitating depression and maanotropic cynicism? Why does Blount's definition feel and read differently than the ones provided by the modern lexicographic works, now used in the Chambers Dictionary or The new Oxford American Dictionary (NOAD)?

coffee /kə'fiːn/ *n.* a powder made by roasting and grinding the seeds of a tree (*Coffea arabica, robusta*, etc.) of the family Rubiaceae; a drink made from the powder; the brown colour of this drink when mixed with milk. [Turk *kahveh*, from *Ar qahwah* orig meaning wine]

Intro to dictionaries



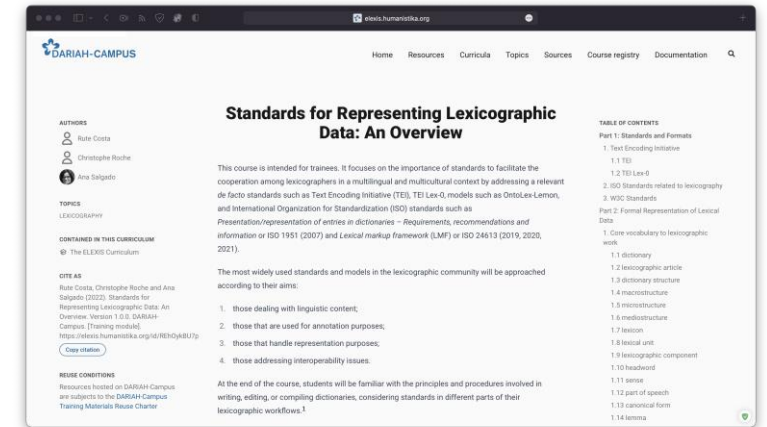
1. The world upside down

Tree structures resemble trees, but in a world turned upside down: the root of the tree is at the top, and the branches are at the bottom. The root element has one or more children. Children with the same parent (there can be only one parent in XML) are siblings etc.

```

graph TD
    entry --> form
    entry --> gramGrp
    entry --> sense
    form --> orth
    gramGrp --> gram
    sense --> def
    sense --> cit
    cit --> quote
    
```

Intro to modeling



Standards for Representing Lexicographic Data: An Overview

This course is intended for trainees. It focuses on the importance of standards to facilitate the cooperation among lexicographers in a multilingual and multicultural context by addressing a relevant de facto standards such as Text Encoding Initiative (TEI), TEI Lex-0 models such as OrotLex-Lemon, and International Organization for Standardization (ISO) standards such as Requirements, recommendations and information or ISO 1951 (2007) and Lexical markup framework (LMF) or ISO 24613 (2019, 2020, 2021).

The most widely used standards and models in the lexicographic community will be approached according to their aims:

- those dealing with linguistic content;
- those that are used for annotation purposes;
- those that handle representation purposes;
- those addressing interoperability purposes.

At the end of the course, students will be familiar with the principles and procedures involved in writing, editing, or compiling dictionaries, considering standards in different parts of their lexicographic workflows.

Lexicographic Standards: An Overview



Lexical Forms

In RDF, it is not possible to describe properties of a string directly. Instead if we wish to provide descriptions of a form, for example saying that **edemata** is a plural form, it is necessary to introduce an identifier for the form that can be described as in the diagram below.

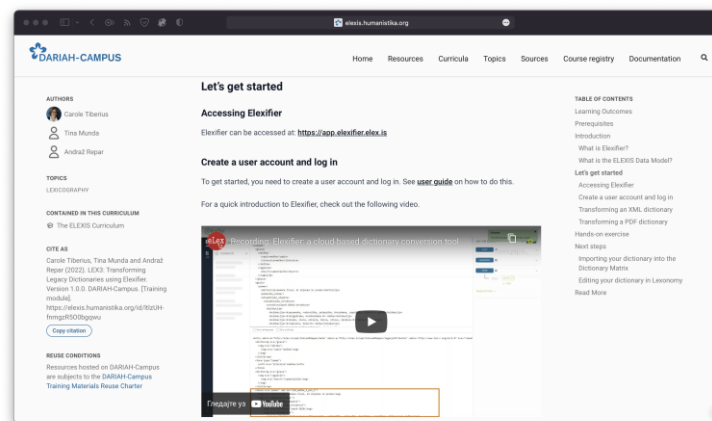
```

graph TD
    Edemata --> numbers-plural
    Edema --> numbers-singular
    numbers-plural --> EDEMA
    numbers-singular --> EDEMA
    
```

Lexical Senses

Similarly, there are often reasons to be able to describe the connection between a lexical entry and a knowledge graph. As such, it is necessary to introduce a named element between the meaning of the word and its lexical entry. We call this element a **lexical sense**, although it should be noted that this term is specific in meaning to this model. A lexical sense is used for properties that are dependent on both a specific meaning and a specific word. For example, definitions are typically given on the knowledge graph elements as they can be shared between synonyms, where as usage examples are typically attached to the lexical sense as they show the usage of a particular word with a particular meaning. In the example, below we indicate that 'slippy' is an antiquated term for the disease. This

Ontolex-Lemon



Let's get started

Accessing Elexifier

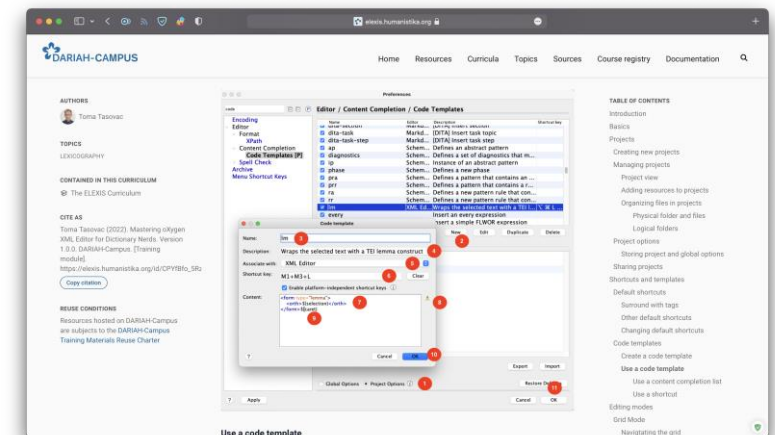
Elexifier can be accessed at <https://app.elexifier.eu>

Create a user account and log in

To get started, you need to create a user account and log in. See [user guide](#) on how to do this. For a quick introduction to Elexifier, check out the following video.

[Elexifier: An Elexifier cloud based dictionary construction tool](#)

Elexifier



Mastering oXygen

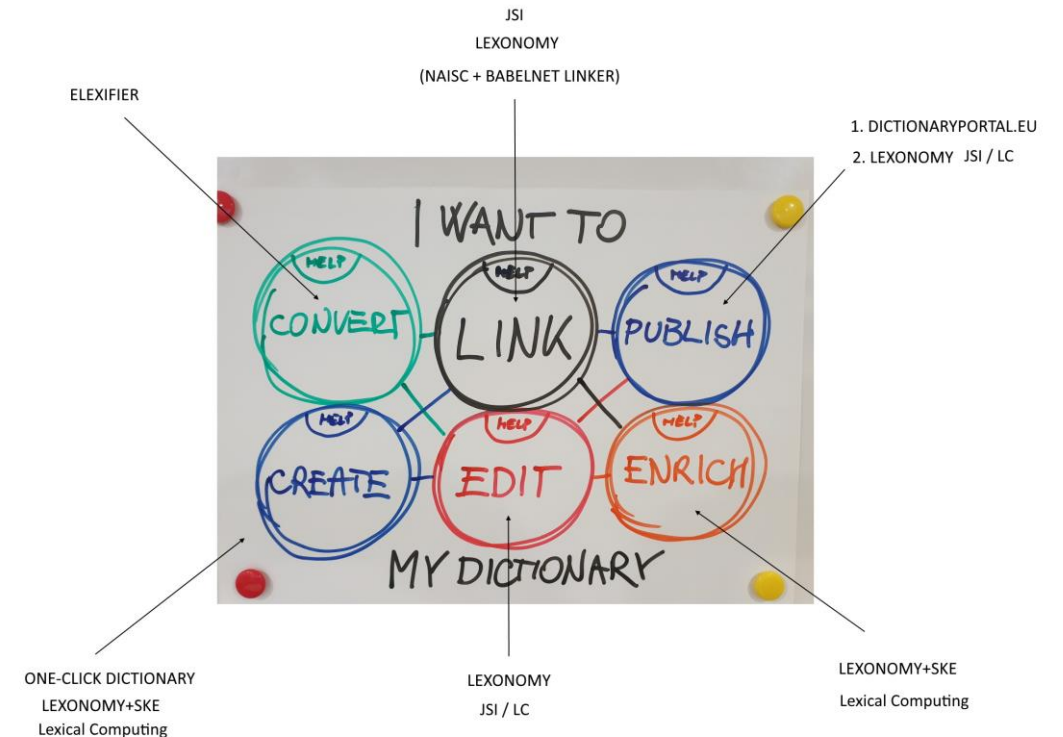
oXygen is a powerful XML editor. This course will guide you through its features and how to use it effectively for lexicographic work.

Mastering oXygen



ELEXIS Curriculum: Multiple Points of Entry

- DARIAH-Campus: integrated curriculum
- ELEXIS: tool-centered





What's next?



What's next?

- Publication of the curriculum on DARIAH-Campus
- Testing the curriculum
- Relax: nothing is perfect
- Updating the curriculum
- Expanding the curriculum
- Ideally, we'll think of the ELEXIS Curriculum as a *living curriculum*, one that can evolve over time, together with this community



Thanks!

